Recurrent-Event Models for Change-Points Detection

Qing Li

Dissertation submitted to the faculty of the
Virginia Polytechnic Institute and State University
in partial fulfilment of the requirements for the degree of

Doctor of Philosophy
in
Statistics

Feng Guo, Chair
Xinwei Deng
Yili Hong
Inyoung Kim

December 16, 2015
Blacksburg, Virginia

Recurrent-Event Models for Change-Points Detection

Qing Li

Abstract

The driving risk of novice teenagers is the highest during the initial period after licensure but decreases rapidly. This dissertation develops recurrent-event change-point models to detect the time when driving risk decreases significantly for novice teenager drivers. The dissertation consists of three major parts: the first part applies recurrent-event change-point models with identical change-points for all subjects; the second part proposes models to allow change-points to vary among drivers by a hierarchical Bayesian finite mixture model; the third part develops a non-parametric Bayesian model with a Dirichlet process prior. In the first part, two recurrent-event change-point models to detect the time of change in driving risks are developed. The models are based on a non-homogeneous Poisson process with piecewise constant intensity functions. It is shown that the change-points only occur at the event times and the maximum likelihood estimators are consistent. The proposed models are applied to the Naturalistic Teenage Driving Study, which continuously recorded *in situ* driving behaviour of 42 novice teenage drivers for the first 18 months after licensure using sophisticated in-vehicle instrumentation. The results indicate that crash and near-crash rate decreases significantly after 73 hours of independent driving after licensure. The models in part one assume identical change-points for all drivers. However, several studies showed that different patterns of risk change over time might exist among the teenagers, which implies that the change-points might not be identical among drivers. In the second part, change-points are allowed to vary among drivers by a hierarchical Bayesian finite mixture model, considering that clusters exist among the teenagers. The prior for mixture proportions is a Dirichlet distribution and a Markov chain Monte Carlo algorithm is developed to sample from the posterior distributions. DIC is used to determine the best number of clusters. Based on the simulation study, the model gives fine results under different scenarios. For the Naturalist Teenage Driving Study data, three clusters exist among the teenagers: the change-points are 52.30, 108.99 and 150.20 hours of driving after first licensure correspondingly for the three clusters; the intensity rates increase for the first cluster while decrease for other two clusters; the change-point of the first cluster is the earliest and the average intensity rate is the highest. In the second part, model selection is conducted to determine the number of clusters. An alternative is the Bayesian non-parametric approach. In the third part, a Dirichlet process mixture model is proposed, where the change-points are assigned a Dirichlet process prior. A Markov chain Monte Carlo algorithm is developed to sample from the posterior distributions. Automatic clustering is expected based on change-points without specifying the number of latent clusters. Based on the Dirichlet process mixture model, three clusters exist among the teenage drivers for the Naturalistic Teenage Driving Study. The change-points of the three clusters are 96.31, 163.83, and 279.19 hours. The results provide critical information for safety education, safety countermeasure development, and Graduated Driver Licensing policy making.

Key Words: Bayesian Finite Mixture Model, Clustering, Constant Piecewise Intensity, Dirichlet Process Mixture Model, Maximum Likelihood Estimate, Naturalistic Teenage Driving Study, Non-Homogeneous Poisson Process.

Dedication

To my Father in heaven, in the pursuit of His glory and honor.

Acknowledgements

I am very grateful to my advisor Feng Guo. He cares the students and is very considerate. He provides much guidance and inspires my interests in Statistics. I could not have found a better mentor. Thank you to Inyoung Kim for offering several great courses and for being a co-author of my manuscripts. Thank you also to Xinwei Deng and Yili Hong for their time and advice as committee members. While at Virginia Tech, I was blessed to be fully sponsored as a Doctoral student. Thank you to Jeffrey Birch for securing this fellowship and for being so encouraging and helpful in my studies. Thank you to the faculty and staff from the Department of Statistics and Virginia Tech for all the support. Thank you to my fellow students, friends, and family. Thank you to my beloved husband for his accompany and support all the time. Finally, thank you to my dear God for being with me, guiding me and taking care of me always.

Contents

List of Figures

List of Tables

Chapter 1 INTRODUCTION

1.1 Background

Traffic crashes are the leading cause of death for teenagers in the United States, and the fatality rate for teenage drivers is substantially higher than that for experienced drivers (National Highway Traffic Safety Administration, 2000). Studies have shown that the driving risk among novice teenagers is the highest early in licensure, declines rapidly for a period of time right after independent driving, and changes slowly later (Williams, 2003). The rapid decline in crash risk after licensure has been documented in a number of recent studies (Mayhew et al., 2003; Lee et al., 2011; Simons-Morton et al., 2011a; Guo et al., 2013). Reducing the high risk in the initial driving period is the foundation for the Graduated Driver Licensing (GDL) law and also the target of teenage driver safety eduction and countermeasure programs.

Various factors contribute to the high driving risk of teenage drivers. Klauer et al. (2014) analyzed the Naturalistic Teenage Driving Study (NTDS) data and concluded that performing secondary tasks while driving would significantly increase the driving risk and the impacts on teenagers were considerably higher than experienced drivers. Jackson et al. (2013) used high G-force events, i.e., hard braking, to predict the occurrence of crash and near-crash (CNC) events and the findings implied that driving style could contribute to the driving risk. Ouimet et al. (2014) showed that teenage drivers with higher cortisol level tend to have lower CNC risk and decrease faster in driving risk over time. Simons-Morton et al. (2014) concluded the driving risk for teenage drivers was directly associated with total time eyes off the forward roadway. The high initial risk and rapid change after licensure are likely to be the results of multiple factors including drivers' experience increase and driving skills.

The duration of the initial high risk period is a key parameter of interest. Many GDL laws impose certain constraints on driving privilege, e.g., teen passengers and night time driving, to decrease the likelihood of CNC events. Using a naturalistic driving study approach, Simons-Morton et al. (2011a) showed the CNC rate declined significantly at six months after licensure. Mayhew et al. (2003) showed similar results with crash database. Guo et al. (2013) examined the variation of the CNC rate change across time and revealed substantial variations among teenage drivers: only about thirty percent drivers at moderate risk showed a significant decrease in CNC rate after six months, while no significant change was found for high and low risk groups.

Driving risk is typically measured by the rate of safety critical events. A common approach for evaluating the time trend of event rate is to aggregate data into pre-defined time intervals, e.g., three-month intervals, for analysis (Simons-Morton et al., 2011a; Guo et al., 2013). The low safety-critical event rate observed in real life requires a long time period as base analysis unit for stable estimations. The long time interval, however, reduces the time resolution and can only detect the time of change in terms of months or even quarters. In addition, the calendar time interval used in many studies does not necessarily correspond well with the actual driving time or mileage, which is arguable to be a more direct measure of driving experience. There is a need for more suitable analytical approaches and new data collection method for estimating the time of risk change.

The Naturalistic Driving Study (NDS) method is a novel driving data collection method characterized by continuous data collection from advanced in-vehicle instrumentation (Dingus et al., 2006; Fitch et al., 2013; Guo et al., 2015b). There are typically no specific instruction for drivers to ensure that the collected data truly reflect their normal drivers behaviour. The NDS provides unrepresented *in situ* driving data under natural, non-experimental driving conditions that could not be obtained from crash database and driving simulator/test track studies. It is especially suitable for evaluating the rapid change in teenage driving risks. The Naturalistic Teenage Driving Study (NTDS) conducted between 2006 and 2009 examined the novice teenage driving. The study included 42 teenage drivers who had just obtained their

Virginia driver licenses; their vehicles were equipped with cameras, accelerometers, Global Positioning Systems, and other vehicle-monitoring devices. The crashes and near-crashes were identified from the collected driving data (Lee et al., 2011). This NTDS provides an opportunity to evaluate the change-point in driving risk for novice teenage drivers.

Change-point models are a natural fit for detecting the risk change-point. Differing from terminal diseases, majority of crashes or safety critical events are not fatal and most likely will occur again in the future. The focus of this dissertation is on developing change-point models in the context of recurrent event. There have been a number of change-point detection models based on constant hazard functions but in a non-recurrent-event context (Matthews and Farewell, 1982; Nguyen et al., 1984; Yao, 1986; Loader, 1991). Ghosh et al. (1993) and Karasoy and Kadilar (2007) presented Bayesian estimators of the change-point assuming a piecewise constant hazard function. Müller and Wang (1994) reviewed different change-point estimators in hazard functions. Pons (2002) estimated the change-point in the time-varying covariates in a Cox model by maximizing the likelihood. Wu et al. (2003) proposed a non-parametric change-point estimator of the hazard function in the counting process context allowing for random censoring. Dupuy (2006) provided consistent estimators of the change-point for both hazard function and regression coefficients in a parametric survival regression model.

There are limited literatures on the change-point detection in the recurrent-event context. Lawless and Zhan (1998) used piecewise constant intensity functions on the analysis of interval-grouped recurrent-event data, where only the event counts in given intervals were recorded. West and Odgen (1997) and Aschar et al. (2007) implemented likelihood approaches to estimate the change-point of the piecewise constant intensity function for only one individual with recurrent events. Frobish and Ebrahimi (2009) proposed a maximum likelihood estimator (MLE) and a Nelson-Aalen estimator for estimating the unknown change-point of multiple individuals with recurrent events.

Recurrent-event models are commonly based on counting processes (Cook and Lawless, 2007, Chap. 1). A counting process is a non-decreasing stochastic process $\{N(t); t \geqslant 0\}$ with positive integer values, in which $N(t)$ is the cumulative number of events till time t. According to the Doob-Meyer decomposition theorem, $N(t) = \Lambda(t) + M(t)$, where $\Lambda(t)$ is the cumulative intensity process and $M(t)$ is a martingale (Andersen et al., 1993). If $\Lambda(t)$ exists, the intensity is $\lambda(t|\pounds_t) = \lim_{\triangledown \to 0} P(N[t, t + \triangledown) \geqslant 1|\pounds_t)/\triangledown$, where $\pounds_t$ is the process history before time t and $\triangledown$ is a small time interval.

The Poisson process is canonical for event count analysis, of which the intensity function $\lambda(t|\pounds_t)$ has the form $\lambda(t)$. Given the marginal probability of an incident at time t, $\lambda(t)$ is the intensity function of a Poisson process. The non-homogeneous Poisson process (NHPP) is a Poisson process when the intensity parameter $\lambda(t)$ is not a constant (Ross, 2006, p. 32). Considering the NHPP $\{N(t); t \geqslant 0\}$ with cumulative intensity $\Lambda(t)$, the number of events that occurred between time zero and t is a random variable that is Poisson distributed with parameter $\Lambda(t)$. The intensity function $\lambda(t)$ of a NHPP can increase or decrease abruptly and the shift point in time is the change-point. All the three parts in this dissertation assume that the event counts are NHPP with constant piecewise intensity functions.

Chapter 2 advances the likelihood-based approach of Frobish and Ebrahimi (2009) and applies two recurrent-event change-point models allowing for multiple unknown change-points. It is shown that the change-points can only occur at the event times and the maximum likelihood estimators (MLE) are consistent. The MLE of the intensity rates are obtained first and the set of event times that maximize the profile likelihood are the MLE of the change-points. The confidence intervals (CIs) and SEs of change-point estimators are obtained by block bootstrapping, where only the drivers are re-sampled with the event times unchanged (Field and Welsh, 2007).

Chapter 2 proposed a recurrent event change-point model assuming identical change-points among novice teenage drivers. The assumption of identical change-points simplifies the analysis, however, studies have shown that risk profiles vary by driver (Guo et al., 2013). Varying change-points by individual driver can accommodate the heterogeneity among drivers, however it would result in over-parametrization issues. It has been demonstrated that clusters existed among teenagers with different patterns of risk change according to Rolls et al. (1991), Ulleberg (2001), Musselwhite (2006), Klauer et al. (2009), and Guo et al. (2013). I propose to use a latent class analysis (LCA) to identify the underlying unknown structure for this study.

LCA can be used to detect subtypes (latent classes) and can classify subjects according to their maximum class membership probability (Lazarsfeld and Henry, 1968; McCutcheon, 1987). The observations in the same cluster are independent in LCA. A commonly used type of LCA is the finite mixture model. In finite mixture models, the density function of observations is a weighted sum of a finite number of known parametric distributions. The weight is the mixture proportion, non-negative and the summation is one. Assuming that the observations are from the mixture distribution, the number of components, the mixture proportions, and the component parameters given the number of components have to be estimated (McLachlan and Peel, 2000).

The inference on finite mixture models can be simplified by Bayesian analysis, which is Bayesian finite mixture model (BFMM) (Diebolt and Robert, 1994). Diebolt and Robert (1994) proposed estimators by Gibbs sampling for a BFMM. Celeux et al. (2000) provided alternatives for Bayesian estimators when multi-modes existed in the posterior. McLachlan and Peel (2000) discussed major issues of BFMM. Richardson and Green (1997), Stephens (2000), and Nobile (2007) presented methods to analyze the BFMM when the number of components was unknown.

The models developed in Chapter 3 are based on BFMM. The drivers are assumed to fall in several clusters. The CNC event count of each teenager follows a NHPP with a piecewise constant intensity function. Li et al. (2015) addressed that the model with one change-point is the best for NTDS. Accordingly, each subject is assumed to have one unknown change-point in this research. The subjects in the same cluster share the identical intensity rates and change-point. The number of clusters, the cluster parameters and the cluster assignments are unknown. Assuming that the change-point of each teenager comes from a finite mixture, the mixture proportions of each subject are assigned a Dirichlet distribution prior in a hierarchical Bayesian framework.

A latent class analysis is conducted in Chapter 3 to address the underlying unknown structure. Traditional latent class modelling contains model selection to choose the best number of clusters. Neal (2000) and Teh (2010) pointed out an alternative was the Bayesian non-parametric approach.

For the Bayesian non-parametric model, the dimension of the parameter space is infinite. To explain a finite sample, the Bayesian non-parametric models (Orbanz, 2010) uses a finite subset of the parameter space adapting to the data. The representations grew with unbounded complexity when observing more data. This type of models are flexible, and can infer the model complexity from the data automatically without conducting Bayesian model comparison. Bayesian non-parametric models are widely used in the field of machine learning, for example, regression, classification, clustering, image segmentation, and natural language processing (Walker et al., 1999; Hjort et al., 2010).

In Bayesian non-parametric models, Dirichlet process (DP) is frequently used, especially in clustering problems when the number of clusters is unknown a priori (Antoniak, 1974; Ferguson, 1983). This type of model is the Dirichlet process mixture model (DPMM), which is a generalization of finite mixture models. DPMM can estimate the number of components and the parameters of the distribution component at the same time (Orbanz, 2010). Section 1.2 introduces DP in detail. Because clusters exist among the teenagers while the number of clusters is unknown, the Bayesian non-parametric model with a DP prior would be promising.

A large part of the literatures on DPMM used MCMC to sample from the posterior distribution. Neal (1992), Escobar (1994), and Bush and MacEachern (1996) developed Gibbs sampling methods for DPMM with conjugate priors. West et al. (1994), Escobar and West (1995), MacEachern and Müller (1998), and Walker and Damien (1998) devised MCMC approaches when using non-conjugate priors. Neal (2000) reviews MCMC methods in DPMM and extended Gibbs sampling to sample from the posteriors. There are other inference approaches like deterministic expectation propagation in Minka and Ghahramani (2003), variational approximation in Blei and Jordan (2006), hierarchical clustering in Heller and Ghahramani (2005), and sequential greedy search in Wang and Dunson (2011).

Chapter 4 proposes a DPMM allowing change-points to vary among drivers. A MCMC algorithm is developed to sample from the posterior distributions. Then the K-means clustering is used to group the subjects. Each subject has one unknown change-point ,and all the subjects share the identical intensity rates. These change-points fall in several clusters while the number of clusters, the cluster parameters and the cluster assignments are unknown. The model is applied to both simulated data and the NTDS data.

1.2 Dirichlet Process

The Dirichlet process (DP) is an infinite-dimensional extension of a Dirichlet distribution (Ferguson, 1973, 1974). The Dirichlet distribution Dirichlet($\alpha_1, \alpha_2, \cdots, \alpha_k$) is a multivariate extension of a Beta distribution, where each parameter is a positive real number (Kotz et al., 2000, Chap. 49). $k > 1$ is the number of categories. The support of a Dirichlet distribution is an open (k-1)-dimensional simplex defined by $\left\{ x_i \in (0,1), \sum_{i=1}^{k} x_i = 1 \right\}$.

A DP has two parameters: a base distribution G_0 and a positive strength parameter α_0. $G \sim DP(\alpha_0, G_0)$ if for any partition $(A_1, A_2, \cdots, A_k)$ of the sample space S,

$$G(A_1), \cdots, (A_k) \sim Dirichlet(\alpha_0 G_0(A_1), \cdots, \alpha_0 G_0(A_k)).$$

The mean distribution of the DP is G_0 which is the best guess of G. The dispersion parameter α_0 shows the strength of the belief that G is G_0 (Escobar, 1994).

If $G \sim DP(\alpha_0, G_0)$, G is a discrete distribution with a countably infinite number of point masses (Blackwell and MacQueen, 1973). Therefore, the probability that two samples will be the same is always non-zero, which means individual values drawn from a DP will likely to repeat and lower values of α_0 causes more repetition. DP is especially useful when the number of clusters is unknown a priori or the number of clusters grows without upper bound as the amount of data increases. For DP, the number of clusters grows in the log scale of the number of observations (Teh et al., 2006).

Various structures can be used to represent a DP, the most important three of which are the Pólya urn scheme, the Chinese restaurant process and the stick-breaking construction.

1. The Pólya urn scheme

   It's also called the Blackwell-MacQueen urn scheme (Blackwell and MacQueen, 1973). It generates a sequence of balls with i.i.d. random colors $\theta_1, \theta_2, \cdots$ from distribution G. The color of the first ball is sampled from G_0. The color of the next ball is either sampled from existing colors or from G_0. Given $m-1$ balls, the conditional distribution of the next ball's color is:

$$\theta_m|\boldsymbol{\theta}_{1:m-1} \sim \frac{\alpha_0 G_0}{\alpha_0 + m - 1} + \frac{\sum_{p=1}^{m-1} \delta(\theta_m, \theta_p)}{\alpha_0 + m - 1}.$$

   where $\delta(\theta_m, \theta_p)$ is a measure defined by

$$\delta(\theta_m, \theta_p) = \begin{cases} 1, \text{ when } \theta_m = \theta_p, \\ 0, \text{ when } \theta_m \neq \theta_p. \end{cases}$$

   This scheme can be utilized to generate a new sample given the existing samples. Furthermore, a new sample has a probability to be drawn from existing samples. Thus the number of distinct values K out of the n draws of θ_p can be smaller than m. This clusters the balls into K groups, while the balls in each group share the same color.

2. The Chinese restaurant process

   As discussed in Aldous (1985), imagine a Chinese restaurant with an infinite number of tables. The first customer will be seated at the first table. A new customer can either sit at an occupied table or a new table.

P(customer m sat at table k|allocation of the previous $m-1$ customers) =

$$\begin{cases} \dfrac{\alpha_0}{\alpha_0+m-1}, & \text{if new table,} \\ \dfrac{n_k}{\alpha_0+m-1}, & \text{otherwise.} \end{cases}$$

n_k is the number of customers at table k. The clustering property of DP is illustrated in the Chinese restaurant process: the more people at a table, the larger chance that a new customer will sit at the table. As a result, only a finite number of tables will be used though there are infinite tables in the restaurant.

3. The Stick-breaking construction

   Sethuraman (1994) showed:

$$G(\theta) = \sum_{k=1}^{\infty} \pi_k \delta(\theta - \theta_k), \text{ where } \theta_k \sim G_0. \tag{1.1}$$

   The weights $\{\pi_k\}_{k=1}^{\infty}$ are constructed by the stick-breaking process, which starts with a stick of unit length. The stick is broken according to the proportion $\beta_k \sim Beta(1, \alpha_0)$. The length of the k^{th} piece is:

$$\pi_k = \beta_k \prod_{l=1}^{k-1} (1 - \beta_l).$$

   $\sum_{k=1}^{\infty} \pi_k = 1$. Stick-breaking process clearly shows that the location of the atoms are sampled from the base distribution G_0 of DP while α_0 affects the weights of these atoms. For smaller α, the first few sticks have significant lengths. For larger α, there will be less variation in the sticks' lengths.

DPs are widely used in the field of statistics and machine learning including: density estimation, mixture models, image processing, and natural language processing (Teh, 2010).

1.3 Naturalist Teenage Driving Study

The NTDS is a naturalistic driving study focusing on novice teenage drivers. The participants consist of 22 females and 20 males with an average age of 16.4 years. The data collection started within two weeks after lincensure. The average driving time over the 18-month study period per participant was 250.63 hours and the stand deviation (SD) is 107.51 hours.

The driving risk is quantitatively measured by crashes and near-crashes rate. A crash involves energy transfer between the participant vehicle and another vehicle or object; a near-crash is similar to a crash except that the driver makes an evasive maneuver to the limit of the vehicle's capacity to avoid a crash (Lee et al., 2011). It is a state-of-practice to combine crashes and near-crashes in driving risk evaluation mainly because of the relatively small sample sizes in naturalistic driving studies (Guo et al., 2010; Klauer et al., 2010; Simons-Morton et al., 2011b; Guo and Fang, 2013; Klauer et al., 2014; Ouimet et al., 2014; Guo et al., 2015a). The validity of using of crash and near-crash has been examined in previous studies (Guo et al., 2010).

A driver might encounter multiple CNC events during the first 18 months after licensure and the average number of CNC events is 6.64 per participant for the NTDS. Fig. 1.1 shows the event plot by driver. The distribution of events is highly unbalanced among drivers: four participants experienced no safety events while 38 involved in a total of 279 CNC events, including 37 crashes and 242 near-crashes. Fig. 1.1 also shows the censoring time which varies from 57.40 to 562.47 hours with a mean of 250.63 and SD of 107.51 hours. The final analysis includes 271 CNC events due to eight missing event times.

The event intervals are based on driving hours. The driving hours, instead of mileage, was used to be consistent with the Graduated Driver Licensing laws, which are typically based on hours of driving. Furthermore, the driving time and distance are highly correlated. Fig. 1.2

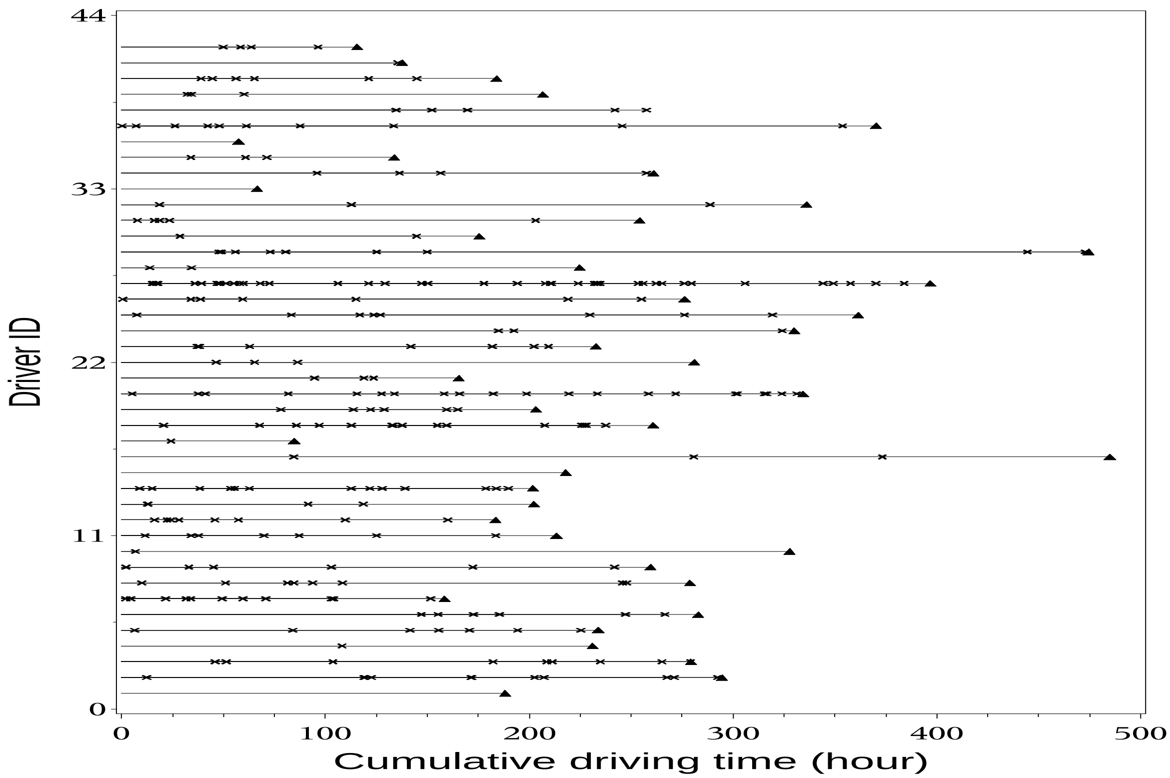

Figure 1.1: Event plot. Cross is the time to event, and triangle is the end of study.

shows the histogram and the boxplot of the inter-event times for all drivers that are heavily skewed to the right. The inter-failure time ranges from 0.0023 to 294.43 hours with a mean of 28.88 and SD of 38.70 hours.

For the drivers with CNC events, Fig. 1.3 shows the histogram and boxplot of the censoring time which varies from 84.70 to 562.47 hours, with a mean of 263.07 and standard deviation (SD) of 102.91 hours. Table 1.1 shows the cumulative average driving time at each month per teenager.

Table 1.1: The cumulative average driving time at each month per teenager

Month	4	6	8	11
Average driving time (hour)	55.07	82.29	109.32	153.79

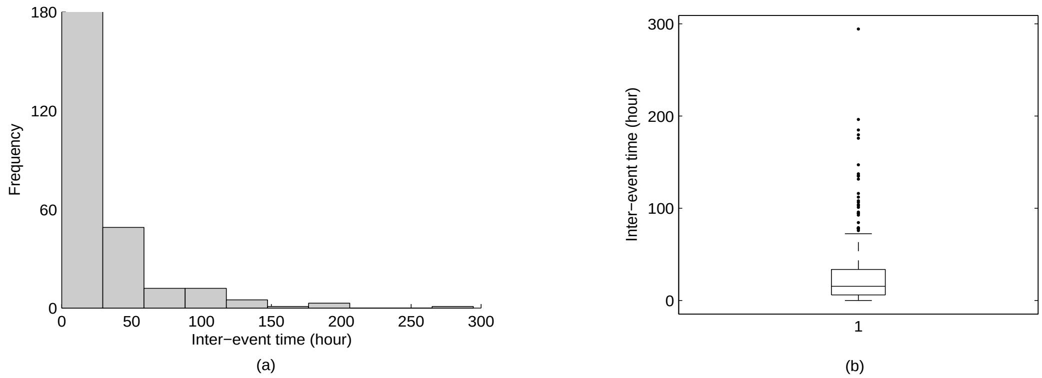

Figure 1.2: Distribution of inter-event time from NTDS data: (a) histogram; (b) boxplot.

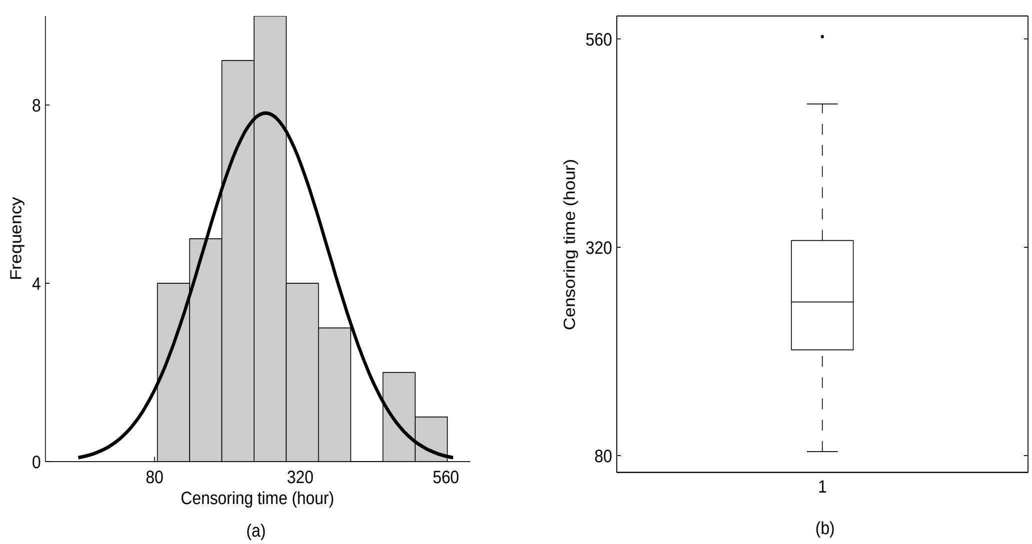

Figure 1.3: The censoring time: (a) histogram; (b) boxplot.

1.4 Dissertation Structure

The rest of this dissertation is organized as follows. Chapter 2 provides the detail for the MLE method to detect the change-points and conducts simulation studies to check the model performance as well as applying the methods to NTDS data. Chapter 3 proposes a hierarchical Bayesian finite mixture model to allow change-points to vary among drivers. In Chapter 4, a Bayesian non-parametric approach where the change-points are assigned a Dirichlet process prior is proposed. Chapter 5 provides the concluding remarks.

Chapter 2 CHANGE-POINTS DETECTION BY MAXIMIZING THE LIKELIHOOD

2.1 Introduction

Driving risk for teenagers in the early period after licensure might be higher but drops fast (Williams, 2003). Instead of change-points in calendar time, the method in this chapter is to detect the change-points in driving time by maximizing the profile likelihood, which is also helpful to understand teenagers' driving risk.

The method advances the likelihood-based approach of Frobish and Ebrahimi (2009) in several ways. First, the constraint that the censoring times be larger than the change-points for all subjects is relaxed. Second, the intensity functions for individuals could vary, which fits the driving risk change better because of high variability among subjects. Third, the standard errors (SEs) of the intensity estimators are inferred by the information matrix in identical-intensity models. The confidence intervals (CIs) and SEs of change-point estimators are obtained by block bootstrapping, where only the drivers are re-sampled with the event times unchanged (Field and Welsh, 2007). Lastly, multiple change-points are allowed in the models. All the change-points are assumed unknown, which is different from the approach in Frobish and Ebrahimi (2009) which considered only two change-points and assumed the first change-point to be known.

The rest of the chapter is as follows. Section 2.2 describes the change-points detection in the context of recurrent-event models. Section 2.3 reports the simulation findings, and Section 2.4 shows the NTDS data application results. Section 2.5 presents concluding remarks.

2.2 Change-Point Models for Recurrent Events

This section presents the identical-intensity model and the subject-specific intensity model to detect when crash and near-crash risks decrease significantly for novice teenage drivers.

The models proposed by Frobish and Ebrahimi (2009) assumed that the event occurrence followed a NHPP with a piecewise constant intensity function having two change-points. Their models require the censoring times to be larger than the second change-point. However, the censoring time varies substantially from 57.40 hours to 562.47 hours for the NTDS. The constraint that the change-point should be smaller than the minimum censoring time would result in an unacceptable upper bound for the change-points. I remove this assumption and the censoring times can be smaller than the change-points.

Furthermore, Rolls et al. (1991), Ulleberg (2001), Dingus et al. (2006), Musselwhite (2006), Klauer et al. (2009), and Guo et al. (2013) have shown that clusters existed among the teenagers with different patterns of risk change. The identical-intensity model assumes that all subjects share the same intensity function and it cannot accommodate the heterogeneity among drivers. To address this issue, a subject-specific intensity model is developed as well.

To develop the notation, denote n_j to be the total number of CNC events and C_j be the cumulative driving time for subject j, $j = 1, \cdots, m$. The events occurred at the ordered times $t_{j1}, \cdots, t_{jn_j}$. The inter-event durations between two consecutive events are $x_{j1}, \cdots, x_{jn_j}$, where $x_{ji} = t_{ji} - t_{j,(i-1)}$, $i = 1, \cdots, n_j$ and $t_{j0} = 0$.

Assume the CNC event counts follow a NHPP with piecewise constant intensity functions and all subjects share the same d unknown change-points, $\boldsymbol{\tau} = (\tau_1, \tau_2, \cdots, \tau_d)$, $1 \leqslant d \leqslant N$ and d is fixed. N is the total number of events for all drivers during the study. $\tau_L \leqslant \tau_1 < \tau_2 < \cdots < \tau_d \leqslant \tau_U$, τ_L and τ_U are the known lower and upper bound of change-points respectively. τ_U is required to guarantee the consistency of the proposed estimator. For the ease of notation in the later formulas, define $\tau_0 = \tau_{d+1} = 0$. The d change-points partition the time axis into $d+1$ intervals: $Y_1, Y_2, \cdots, Y_{d+1}$, $Y_p = (\tau_{p-1}, \tau_p]$ for $p = 1, 2, \cdots, d$, $Y_{d+1} = (\tau_d, +\infty)$. S_p is the index set where $C_j \in Y_p$. The number of elements in S_p is k_p. $N_j^{(p)}$ is the number of events for the j^{th} driver in Y_p, $\sum_{p=1}^{d+1} N_j^{(p)} = n_j$. I_p is the indicator function on Y_p, $p = 1, 2, \cdots, d+1$. Each driver joined the study at time zero. Depending on whether the intensity functions are treated as equal among subjects, two models are proposed, namely the identical-intensity model and the subject-specific intensity model as discussed below.

2.2.1 Identical-intensity model

This model generalizes the likelihood-based model by Frobish and Ebrahimi (2009) to allow for d unknown change-points. I also relax the restriction that the change-points should be smaller than the censoring times, as some participants might drive very short period of time.

Assume that m drivers share the identical piecewise constant intensity function $\lambda(t) = \sum_{p=1}^{d+1} \lambda_p I_p(t)$. Integrating $\lambda(t)$ yields

$$\Lambda(t) = \sum_{p=1}^{d+1} \left[\sum_{q=1}^{p-1} \lambda_q (\tau_q - \tau_{q-1}) + \lambda_p (t - \tau_{q-1}) \right] I_p(t).$$

The likelihood for the jth driver (Thompson, 1988) is

$$L_j(\boldsymbol{\lambda}, \boldsymbol{\tau}) = exp[-\Lambda(C_j)] \prod_{i=1}^{n_j} \lambda(t_{ji}) = exp[-\Lambda(C_j)] \prod_{p=1}^{d+1} \lambda_p^{N_j^{(p)}}, \ \boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_{d+1}).$$

The log-likelihood for all the drivers combined is

$$LogL(\boldsymbol{\tau}, \boldsymbol{\lambda}) = \sum_{p=1}^{d+1} \log \lambda_p \sum_{j=1}^{m} N_j^{(p)} - \sum_{j=1}^{m} \Lambda(C_j),$$

where $\sum_{j=1}^{m} \Lambda(C_j) = \sum_{p=1}^{d+1} \lambda_p \sum_{j \in S_p} C_j + \sum_{p=1}^{d} (m - \sum_{q=1}^{p} k_q) \tau_p (\lambda_p - \lambda_{p+1})$.

Setting the derivatives of $logL(\cdot)$ over $\boldsymbol{\lambda}$ to $\mathbf{0}$ and solving equations yields the MLE:

$$\widehat{\lambda}_p = \frac{\sum\limits_{j=1}^{m} N_j^{(p)}}{\sum\limits_{j\in S_p} C_j - \tau_{p-1}(m - \sum\limits_{q=1}^{p-1} k_q) + \tau_p(m - \sum\limits_{q=1}^{p} k_q)}, \ p = 1, 2, \cdots, d+1. \tag{2.1}$$

$\dfrac{\partial^2 logL(\boldsymbol{\tau}, \boldsymbol{\lambda})}{\partial \lambda_p^2} = -\dfrac{\sum\limits_{j=1}^{m} N_j^{(p)}}{\lambda_p^2}$, which are always non-positive, indicating that $\widehat{\lambda}_p$'s in Eq. 2.1 are the MLE. $\boldsymbol{\tau}$ has to be estimated first to obtain the numerical value of $\widehat{\boldsymbol{\lambda}}$.

The standard error (SE) of the estimation can be derived from the information matrix (Lehmann and Casella, 1998),

$$SE(\widehat{\lambda}_p) = \frac{\widehat{\lambda}_p}{\sqrt{\sum\limits_{j=1}^{m} N_j^{(p)}}}, \ p = 1, \cdots, d+1. \tag{2.2}$$

Following Frobish and Ebrahimi (2009), I derive the following theorems.

Theorem 1. The values of $\widehat{\tau}_1, \widehat{\tau}_2, \cdots, \widehat{\tau}_d$ that maximize $logL(\widehat{\boldsymbol{\lambda}}, \boldsymbol{\tau})$ in the identical-intensity model can only locate at the observed event times.

Proof. Plug $\widehat{\boldsymbol{\lambda}}$ into Eq. 2.2.1, the profile log-likelihood is:

$$logL(\widehat{\boldsymbol{\lambda}}, \boldsymbol{\tau}) = \sum_{p=1}^{d+1} \log \frac{\sum\limits_{j=1}^{m} N_j^{(p)}}{\sum\limits_{j\in S_p} C_j - \tau_{p-1}(m - \sum\limits_{q=1}^{p-1} k_q) + \tau_p(m - \sum\limits_{q=1}^{p} k_q)} \sum_{j=1}^{m} N_j^{(p)} - N.$$

$logL(\widehat{\boldsymbol{\lambda}}, \boldsymbol{\tau})$ is not a continuous function of $\boldsymbol{\tau}$ but leaps at every event time. Order $\{t_{ji} | 1 \leq j \leq m, 1 \leq i \leq n_j\}$ to be $t_{(1)}, t_{(2)}, \cdots, t_{(N)}$. For $p = 1, 2, \cdots, d$, when $t_{(c_p)} \leqslant \tau_p \leqslant t_{(c_p+1)}$ for any fixed index c_p, $N_j^{(p)}$, $\sum\limits_{j\in S_p} C_j$ and $\sum\limits_{q=1}^{p-1} k_q$ are also fixed. Consequently, $logL(\widehat{\boldsymbol{\lambda}}, \boldsymbol{\tau})$ is continuous and differentiable in any fixed set $\left\{t_{(c_p)} \leqslant \tau_p \leqslant t_{(c_p+1)},\ p=1,2,\cdots,d\right\}$. It follows that

$$\frac{\partial^2 logL(\cdot)}{\partial \tau_p^2} = \frac{(m-\sum\limits_{q=1}^{p} k_q)^2 \sum_{j=1}^{m} N_j^{(p)}}{\left\{\sum\limits_{j\in S_p} C_j - \tau_{p-1}(m-\sum\limits_{q=1}^{p-1} k_q) + \tau_p(m-\sum\limits_{q=1}^{p} k_q)\right\}^2} + \frac{(m-\sum\limits_{q=1}^{p} k_q)^2 \sum_{j=1}^{m} N_j^{(p+1)}}{\left\{\sum\limits_{j\in S_{p+1}} C_j - \tau_p(m-\sum\limits_{q=1}^{p} k_q) + \tau_{p+1}(m-\sum\limits_{q=1}^{p+1} k_q)\right\}^2},$$

which is non-negative. Thus, the values that maximize $logL(\cdot)$ on each interval $[t_{(c_p)}, t_{(c_p+1)}]$ can only be one of the endpoints which are the event times. $\widehat{\boldsymbol{\tau}}$ is the set of d event times between τ_L and τ_U that maximizes $logL(\cdot)$:

$$\widehat{\boldsymbol{\tau}} = \underset{\tau_L\leqslant t_p\leqslant\tau_U, t_p=t_{ji}|1\leq p\leq d, 1\leq j\leq m, 1\leq i\leq n_j}{\operatorname{argmax}} logL(\boldsymbol{t}),\ \boldsymbol{t}=(t_1,t_2,\cdots,t_d). \tag{2.3}$$

□

Theorem 2. The MLE of $\boldsymbol{\tau}$ on $[\tau_L, \tau_U]$ for the identical-intensity model is consistent.

Proof. Following Frobish and Ebrahimi (2009), $logL(\cdot)$ is finite on $[\tau_L, \tau_U]$, which satisfies the conditions listed in Van Der Vaart (1998, p.48, Theorem 5.14). The MLE of $\boldsymbol{\tau}$ on $[\tau_L, \tau_U]$ is consistent as a result. □

2.2.2 Subject-specific intensity model

Considering a change-point model where intensity function varies by subject, the intensity for subject j is $\lambda_j(t) = \sum\limits_{p=1}^{d+1} \lambda_{jp} I_p(t),\ j=1,\ \cdots,\ m$. Following the similar steps as in Section 2.2.1, $logL(\cdot) = \sum\limits_{p=1}^{d+1}\sum\limits_{j=1}^{m} N_j^{(p)} \log\lambda_{jp} - \sum\limits_{j=1}^{m} \Lambda(C_j)$, where $\sum_{j=1}^{m} \Lambda(C_j) = \sum\limits_{p=1}^{d+1}\sum\limits_{j\in S_p} C_j\lambda_{jp} +$

$\sum_{p=1}^{d} \tau_p \sum_{j \in (\cup_{q=1}^{p} S_q)^c} (\lambda_{jp} - \lambda_{j,p+1})$. The MLE of λ_p's are:

$$\widehat{\lambda}_{jp} = \begin{cases} 0, j \in \cup_{q=1}^{p-1} S_q, \\ \dfrac{N_j^{(p)}}{C_j - \tau_{p-1}}, j \in S_p, \\ \dfrac{N_j^{(p)}}{\tau_p - \tau_{p-1}}, j \in (\cup_{q=1}^{p} S_q)^c, \end{cases} \qquad p = 1, 2, \cdots, d+1. \tag{2.4}$$

Following similar argument as in Section 2.2.1, I have the following theorems. *Theorem 3.* The values of $\widehat{\tau}_1, \widehat{\tau}_2, \cdots, \widehat{\tau}_d$ that maximize $logL(\cdot)$ in the subject-specific intensity model locate at the event times.

Proof. $LogL(\cdot) = \sum_{p=1}^{d} \sum_{j \in (\cup_{q=1}^{p} S_q)^c} N_j^{(p)} \log \dfrac{N_j^{(p)}}{\tau_p - \tau_{p-1}} + \sum_{p=1}^{d+1} \sum_{j \in S_p} N_j^{(p)} \log \dfrac{N_j^{(p)}}{C_j - \tau_{p-1}} - N$. On each interval $[t_{(c_p)}, t_{(c_p+1)}]$,

$$\frac{\partial^2 logL(\cdot)}{\partial \tau_p^2} = \frac{\sum_{j \in (\cup_{q=1}^{p} S_q)^c} N_j^{(p)}}{(\tau_p - \tau_{p-1})^2} + \sum_{j \in S_{p+1}} \frac{N_j^{(p+1)}}{(C_j - \tau_p)^2} + \frac{\sum_{j \in (\cup_{q=1}^{p+1} S_q)^c} N_j^{(p+1)}}{(\tau_{p+1} - \tau_p)^2}, \ p = 1, \ \cdots, \ d,$$

which are non-negative for all values of τ_p. $\square$

Theorem 4. The MLE of $\boldsymbol{\tau}$ on $[\tau_L, \tau_U]$ for the subject-specific intensity model is consistent.

Proof. Similarly to Frobish and Ebrahimi (2009), $logL(\cdot)$ is finite on $[\tau_L, \tau_U]$, which guarantees the consistency of the MLE of $\boldsymbol{\tau}$ on $[\tau_L, \tau_U]$. $\square$

Note: for all the models in this chapter, $\sum_{j=1}^{m} \widehat{\Lambda}(C_j) = N$. This is consistent with the fact that the other component of the counting process $N(t)$ is a martingale, which has a mean of zero.

2.3 Simulation Studies

A simulation study is conducted to investigate the performance of the proposed models with different sample sizes, intensity rates, and change-point values.

Data from a NHPP with piecewise constant intensity functions is generated based on the distribution of the inter-event times (Klein and Roberts, 1984). For each subject, the cumulative density function (CDF) of the i^{th} inter-event time $X_i = T_i - T_{(i-1)}$ conditional on the previous event times $T_0 = t_1 = 0,\ T_1 = t_1,\ \cdots,\ T_{i-1} = t_{i-1}$ is

$$F_{t_{i-1}}(x) = Pr\left[X_i \leq x | T_s = t_s, s = 1, 2, \cdots, i-1\right] = 1 - exp\left[\Lambda(t_{i-1}) - \Lambda(t_{i-1} + x)\right].$$

Given the first i events, the next event time t_{i+1} is the i^{th} event time plus the $(i+1)^{th}$ inter-event time with CDF F_{t_i}. Accordingly, the generation of the data follows the procedure below:

1. $i = 0$, initialize $t_{(i)} = 0$;
2. Generate $x_{(i+1)}$ with CDF F_{t_i} randomly;
3. $t_{(i+1)} = t_{(i)} + x_{(i+1)}$;
4. $i = i + 1$, go to step (2).

For the j^{th} individual, repeat until t_{i+1} is larger than the censoring time C_j. $t_1,\ t_2,\ \cdots,\ t_i$ are the event times. Run the process m times to generate event times for m drivers.

For each combination of parameters, I generate $B = 5,000$ data sets with censoring times uniformly from 400 to 500. For the subject-specific intensity model with one change-point, assume $\lambda_{j1} \sim Exponential\ (40)$ and $\lambda_{j2} \sim Exponential\ (20)$, where 40 and 20 are the means of the exponential distributions. For the subject-specific intensity model with two change-points, assume $\lambda_{j1} \sim Exponential\ (10)$, $\lambda_{j2} \sim Exponential\ (40)$, and $\lambda_{j3} \sim Exponential\ (10)$.

Models in Section 2.2 are applied to the simulated data. τ_L is 0 and τ_U is 300. The root mean square error (RMSE) of $\widehat{\tau}$ is $\sqrt{(1/B)\sum_{k=1}^{B}(\widehat{\tau}-\tau)^2}$ and absolute percentage bias (%) is $(1/B)|\sum_{k=1}^{B}(\widehat{\tau}-\tau)|/\tau \times 100\%$. The coverage probabilities of the 95% CI are obtained by bootstrapping 1,000 times.

Table 2.1 shows the average, RMSE, absolute percentage bias (%) and coverage probability for the identical-intensity model. The RMSEs are reasonally small and the absolute percentage biases (%) of the parameters are within 3.7% of the true values. The coverage probabilities are above 86%. When the sample size m increases, the RMSE, absolute percentage bias (%) and the coverage probability tend to decrease. The estimators behave well overall in the identical-intensity model.

Table 2.2 shows the results of the subject-specific intensity model with one or two change-points. When the sample size m increases, the RMSE, absolute percentage bias (%) and the coverage probability become smaller. $\widehat{\boldsymbol{\tau}}$ behaves fine for the subject-specific intensity model.

In summary, the RMSE, absolute percentage bias (%) and the coverage probability decrease when the number of drivers increases. For the subject-specific models, RMSE and |bias (%)| of the parameters are greater than the identical-intensity model, potentially due to increased model complexity and the limited information from the data. Overall, these models give satisfactory results under different scenarios.

2.4 Application of the Intensity Models to the NTDS

I applied both types of models with no more than five change-points to the NTDS data. Because of the relatively small sample size, fitting models with more change-points would result in instability of the estimators. Zero and 300 hours are chosen as the lower and upper bound of change-points respectively. The SEs of rates are estimated by Eq.2.2 for the identical-intensity model and the SEs of the change-points for both types of models are estimated by block bootstrapping. CIs are estimated by block bootstrapping, where only the drivers are re-

Table 2.1: Simulation results for the identical-intensity model

No. of change-points	m	Parameter	True value	Average of estimates	RMSE	|Bias (%)|	Coverage probability (%)
1	40	τ	60	59.10	5.2	1.5	86.8
		λ_1	30	30.76	3.8	2.5	94.1
		λ_2	10	9.95	0.8	0.5	93.8
	80	τ	60	59.55	3.1	0.7	88.2
		λ_1	30	30.40	2.6	1.3	95.3
		λ_2	10	9.98	0.6	0.2	94.2
	40	τ	90	89.00	5.1	1.1	86.6
		λ_1	30	30.52	3.0	1.7	94.5
		λ_2	10	9.93	0.8	0.7	93.5
	80	τ	90	89.50	3.0	0.6	87.7
		λ_1	30	30.25	2.1	0.8	95.0
		λ_2	10	9.96	0.6	0.4	94.2
2	40	τ_1	50	51.03	3.8	2.1	86.6
		τ_2	120	119.29	9.9	0.6	94.0
		λ_1	10	10.18	2.3	1.8	96.0
		λ_2	40	41.49	4.7	3.7	93.2
		λ_3	20	19.91	1.3	0.5	93.3
	80	τ_1	50	50.35	1.6	0.7	87.0
		τ_2	120	119.59	5.1	0.3	93.7
		λ_1	10	10.07	1.6	0.7	94.9
		λ_2	40	40.56	2.9	1.4	95.5
		λ_3	20	19.98	0.9	0.1	94.0
3	40	τ_1	80	79.61	4.8	0.5	91.5
		τ_2	150	150.51	4.9	0.3	89.8
		τ_3	220	219.29	4.8	0.3	88.1
		λ_1	60	60.69	4.4	1.1	98.3
		λ_2	30	29.08	3.5	3.1	93.2
		λ_3	60	61.45	5.2	2.4	96.6
		λ_4	30	29.75	2.1	0.8	94.9

Table 2.2: Estimation of τ in simulation for the subject-specific intensity model

No. of change-points	m	τ	Average of estimates	RMSE	|Bias (%)|	Coverage probability (%)
1	40	60	68.01	19.8	13.4	89.7
	80	60	65.19	14.9	8.6	93.1
	120	60	65.14	14.7	8.6	94.6
	40	75	79.07	19.6	5.4	87.0
	80	75	76.40	13.6	1.9	90.4
	120	75	76.38	13.6	1.8	93.0
	40	90	90.75	19.9	0.8	85.6
	80	90	87.60	13.2	2.7	86.2
	120	90	88.83	12.9	1.3	88.5
2	40	$\tau_1 = 50$	53.36	7.2	6.7	83.7
		$\tau_2 = 120$	114.16	10.4	4.9	87.7
	80	$\tau_1 = 50$	50.50	5.6	7.0	89.6
		$\tau_2 = 120$	115.29	8.2	3.9	89.4

sampled 10,000 times and the event-times remain unchanged. The best model is chosen based on the Akaike information criterion (AIC), which combines the goodness of fit for the data and a penalty for model complexity (Akaike, 1974). A smaller AIC indicates a better model and AIC difference greater than four suggests a significant difference between two models (Burnham and Anderson, 2002).

For the identical-intensity model with one change-point, Fig. 2.1 illustrates the change of log-likelihood over change-point and the estimated cumulative function for all drivers. The slopes of the cumulative functions are the corresponding intensity rates, which are the average number of CNC events per teenager per 1,000 hours. Table 2.3 shows that the change-point is at 65.29 hours of driving after first licensure. The intensity before the change-point is 33.65 CNC events per teenager per 1,000 hours and becomes 22.97 after the change-point, which is equivalent to the rates computed from the NTDS data directly by the following equation:

$$Average\ incident\ rate = \frac{Total\ number\ of\ events}{Total\ driving\ time \times Number\ of\ drivers}.$$

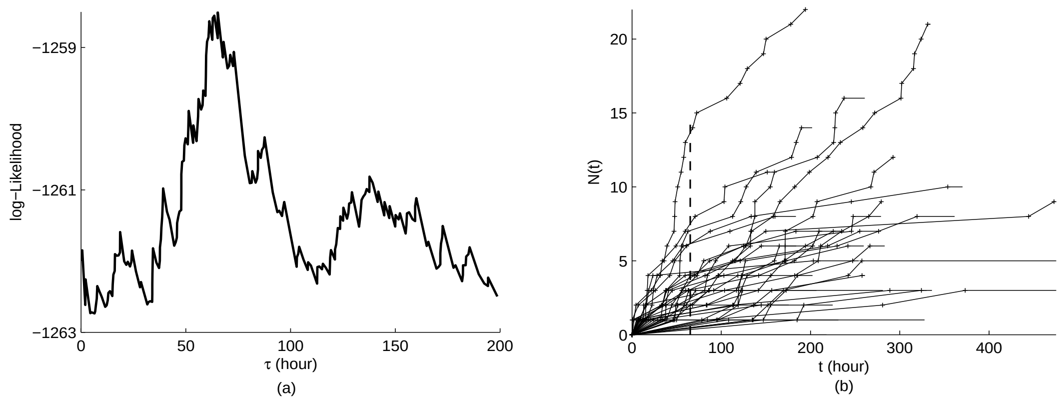

Figure 2.1: The identical-intensity model with one change-point: (a) log-likelihood vs. change-point τ, the highest log-likelihood corresponds to the MLE of change-point; (b) the cumulative event counts of each driver over driving time t. The counts are marked by crosses and connected by drivers. The censoring times are connected with the last event times by horizontal lines for each driver. The vertical dashed-line shows the estimated τ.

For the identical-intensity model with two change-points, τ_2 is close to the change-point estimated by the identical-intensity model with one change-point. Table 2.3 shows that the intensity increases from 28.02 CNC events per teenager per 1,000 hours to 41.22 and later drops to 23.12.

For the subject-specific intensity model with one change-point, Fig. 2.2 shows the log-likelihood over change-point and the estimated cumulative intensity function for each driver. Table 2.3 shows that the change-point is at 72.95 hours of driving after first licensure. Figure 2.3 shows two different patterns: approximately half of the drivers have lower risk after certain time of driving while the risk for others increases. $\widehat{\lambda}_1$ is larger than $\widehat{\lambda}_2$ on the average.

For the subject-specific model with two change-points, Table 2.3 shows the second change-point is equal to the change-point estimated by the subject-specific model with one change-point. The first change-point is close to the first change-point for the identical-intensity model with two change-points. Fig. 2.4 displays the histogram of the rates for all drivers with the

Table 2.3: The NTDS application results

Model	No. of change-points	Parameter	Estimate	95% CI	SE	AIC
	1	τ	65.29	[42.29, 95.96]	10.86	
		λ_1	33.65	[21.72, 47.69]	3.51	2523.0
		λ_2	22.97	[15.30, 31.91]	1.72	
Identical-intensity	2	τ_1	33.99	[24.32, 72.95]	16.40	
		τ_2	61.23	[58.36, 128.94]	24.33	
		λ_1	28.02	[18.03, 45.91]	4.43	2524.0
		λ_2	41.22	[9.21, 60.16]	6.01	
		λ_3	23.12	[14.16, 34.71]	1.70	
Subject-specific intensity	1	τ	72.95	[45.78, 93.76]	11.47	2415.2
	2	τ_1	31.65	[31.65, 72.56]	20.09	2446.6
		τ_2	72.95	[62.71, 108.41]	22.20	

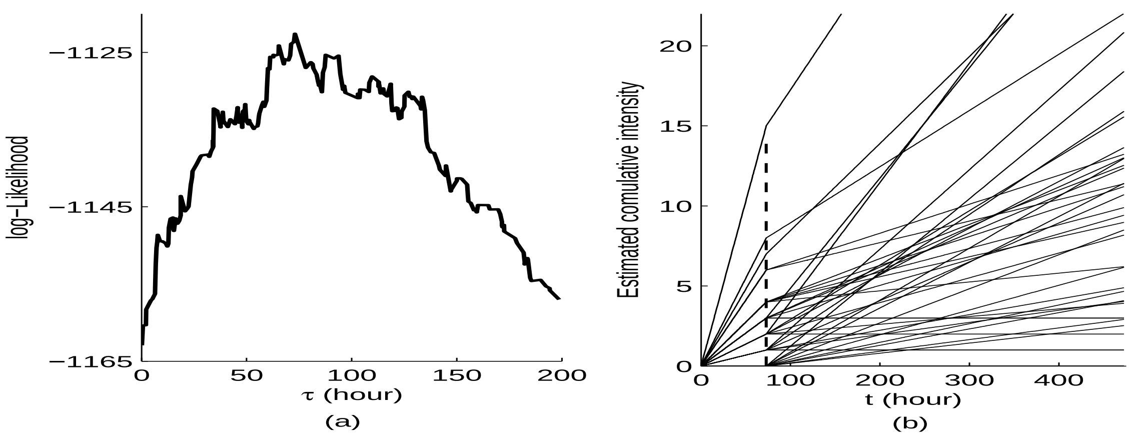

Figure 2.2: The subject-specific intensity model with one change-point: (a) log-likelihood vs. change-point τ; (b) the estimated cumulative intensity function $\Lambda_j(t)$ for each driver with the estimated change-point in the vertical line.

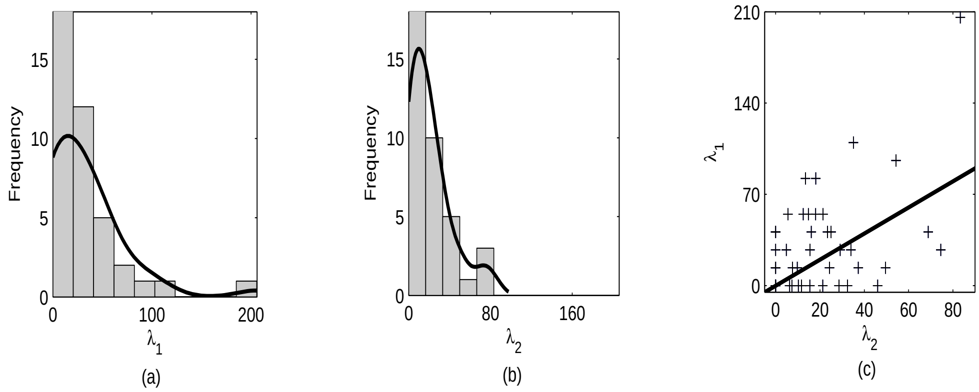

Figure 2.3: The intensity rates for all drivers of the subject-specific intensity model with one change-point: (a) histogram of $\widehat{\lambda}_1$ with kernel fitting; (b) histogram of $\widehat{\lambda}_2$ with kernel fitting; (c) scatter plot of $\widehat{\lambda}_1$ vs. $\widehat{\lambda}_2$.

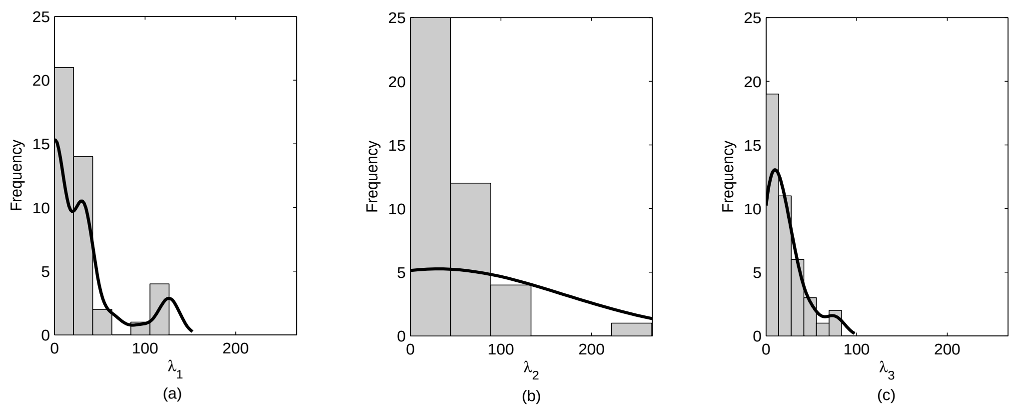

Figure 2.4: The histogram and kernel fitting of intensity rates for all drivers in the subject-specific intensity model with two change-points: (a) $\widehat{\lambda}_1$; (b) $\widehat{\lambda}_2$; (c) $\widehat{\lambda}_3$.

Table 2.4: Intensity rates of the subject-specific intensity model

No. of change-points	Parameter	Average of λ	95% CI	SD	NTDS average
1	λ_1	32.31	[21.22, 46.35]	39.57	32.55
	λ_2	20.84	[15.24, 27.92]	21.05	22.98
2	λ_1	27.84	[19.71, 46.48]	6.93	27.84
	λ_2	35.74	[14.11, 53.37]	11.09	36.20
	λ_3	20.84	[14.43, 35.99]	5.07	22.98

kernel fitting.

For the subject-specific intensity model, the sample averages of the rates are close to the NTDS averages as shown in Table 2.4, which indicates a decent model fitting.

In brief, the subject-specific intensity model with one change-point has the smallest AIC, hence the best model. According to this model, approximately half of the drivers have lower risk after 72.95 hours of driving while the risk for others increases. On the average, the intensity rate decreases from 32.31 CNC events per teenager per 10,000 hours to 20.84. The AIC is increasing when the number of change-points is increasing for both types of models.

2.5 Conclusion and Discussion

Crash rates among novice teenage drivers are high initially, yet decline rapidly for a period of time, and then decline relatively slower thereafter. The observation is consistent with the development of expertise in complex psycho-motor tasks of all sorts (Keating, 1978). It is useful to identify the change-points in crash rates which may quantify the amount of experience required for relatively safe driving.

Novice drivers make more mistakes with driving initially and their driving improves with experience: rapidly at first but more slowly thereafter. It is critical to identify the period during which novices are at greater risk so that parents and policy makers can provide more guidance to protect them and other drivers from their mistakes. Previous research suggested that the change-point occurred at about six months of driving. However, the actual amount of driving time varies considerably during the first six months after licensure. Therefore, identifying the change-point of risk in driving hours provides critical information to improve young drivers' education, safety counter measures, and Graduated Driver Licensing regulations.

This chapter develops two alternative recurrent-event models for detecting the change-point of driving risk for novice teenage drivers. The analysis shows that the subject-specific intensity model with one change-point gives the best fit among the models for NTDS: the change-point is at 72.95 hours of driving after licensure with a 95% CI [45.78, 93.76]. The average driving time per teenager in the NTDS was 82.29 hours for six months and 68.12 hours for five months. Therefore, the change-point obtained by the subject-specific intensity model with one change-point is between five and six months after licensure for most novice teens, which is consistent with the finding of Mayhew et al. (2003). In practice, the subject-specific intensity model is more flexible in accommodating different risk patterns among subjects because of variability in CNC rates and change-points in the sample data.

There are several future extensions for the current work. Though the best model for NTDS is the subject-specific intensity model, this type of model would result in over-parametrization, especially when considering multiple change-points. An alternative is the mixed effect model with random intensity rates. Another possible extension would be relaxing the condition on identical change-points for all drivers, because clusters may exist among the teenagers considering the different patterns of change Rolls et al. (1991). Specifically, interest lies in determining if there are particular subgroups with different underlying change patterns. Other possible extensions would be considering covariates like gender to predict the change-points and changing the form of the piecewise constant intensity function. The resulting findings would help in developing provisions on graduated licensing policies for teenage drivers to encourage low-risk driving (Williams, 2003).

Chapter 3 CHANGE-POINTS DETECTION IN DRIVING RISK BY A HIERARCHICAL BAYESIAN FINITE MIXTURE MODEL

3.1 Introduction

The leading cause of mortality in teenagers and young adults is motor vehicle crashes (Centers for Disease Control and Prevention, 2008). A wide variety of studies are conducted on the driving risk of teenage drivers (Deery and Fildes, 1999). Williams (2003) showed that the driving risk of novice teenage drivers decreased significantly in the early period after licensure and became stable later. Mayhew et al. (2003), Lee et al. (2011), Simons-Morton et al. (2011a), and Guo et al. (2013) showed that the driving risk changed around six months after licensure. The point when the driving risk pattern changes is the change-point.

One common issue is that most existing research aggregated data into pre-defined time intervals for analysis, e.g., three-month intervals, which reduced the time resolution. Models without data aggregation will be necessary. Another issue is that limited research has been conducted to evaluate the change-point of driving risk in driving hours. The change-point in driving time provides an important measure on how much experience is required before relatively safer driving occurs (Simons-Morton et al., 2011a). The Graduated Driver Licensing or the learner's permit laws for teenagers are also based on driving time.

Chapter 2 and Li et al. (2015) assumed the identical change-points among novice teenage drivers. The assumption of identical change-points simplifies the analysis, however, different drivers have different risk patterns and are likely to have different change-points (Guo et al., 2013). Varying change-points are necessary to accommodate the heterogeneity among drivers. Yet, it would result in over-parametrization if each driver has a change-point parameter.

According to Rolls et al. (1991), Ulleberg (2001), Musselwhite (2006), and Klauer et al. (2009), clusters existed among teenagers with different patterns of risk change. A latent class analysis (LCA) is appropriate to identify the underlying unknown structure.

LCA is a type of analysis to detect latent classes and to classify subjects according to their maximum class membership probability (Lazarsfeld and Henry, 1968; McCutcheon, 1987). The Bayesian finite mixture model (BFMM) is commonly used in LCA. Diebolt and Robert (1994) proposed estimators by Gibbs sampling for a Bayesian finite mixture model. Celeux et al. (2000) provided alternatives for Bayesian estimators when multi-modes existed in the posterior. McLachlan and Peel (2000) discussed major issues of BFMM. Richardson and Green (1997), Stephens (2000), and Nobile (2007) presented different methods to analyze the BFMM when the number of components was unknown.

This chapter proposes BFMM. The drivers are assumed to fall in several clusters. The CNC event count of each teenager follows a non-homogeneous Poisson process (NHPP) with a piecewise constant intensity function. Each subject has one unknown change-point while the subjects in the same cluster share the identical intensity rates and change-point. These change-points fall in several clusters: the number of clusters, the cluster parameters and the cluster assignments are unknown. Assuming that the change-point of each teenager comes from a finite mixture, the mixture proportions of each subject are assigned a Dirichlet distribution prior in a hierarchical Bayesian framework.

The convergence of the Markov chain Monte Carlo (MCMC) is examined by the Gelman-Rubin approach (Gelman and Rubin, 1992). Without knowing how many clusters exist among the teenage drivers, the models with different number of clusters are compared. The model with the smallest DIC (Spiegelhalter et al., 2002) will be the best model. Each subject is assigned to the cluster with the largest class membership probability. I check the model performance under different parameter settings and apply the methodology to the NTDS data.

The rest of the chapter includes the following sections. Section 3.2 develops the BFMM to classify the teenage drivers based on their change-points. Section 3.3 applies the method to the simulated data and checks the sensitivity of the method in different scenarios. Section 3.4 applies the model to the NTDS data. Section 3.5 presents the conclusion and discusses the future work.

3.2 The Hierarchical Bayesian Finite Mixture Model for Change-Point Detection

Li et al. (2015) proposed change-point detection models assuming that all the drivers share the same change-points. However, the patterns of risk change vary among the teenagers and subgroups exist (Guo et al., 2013). The models with identical change-points assumption cannot address the heterogeneity among drivers. Correspondingly, I develop a hierarchical model based on BFMM where the change-points of the drivers fall in several clusters.

3.2.1 The finite mixture model for change-point detection

There is at least one CNC event per teenager and the total number of events for the j^{th} driver is n_j. The ordered times of these events are $t_{j1},\ t_{j2},\ \cdots,\ t_{jn_j}$. The total number of events for all drivers during the study period is $N = \sum_{j=1}^{m} n_j$. The total driving time for the j^{th} subject is C_j during the study, $j = 1,\ \cdots,\ m$, and m is the total number of drivers.

Considering the NHPP with cumulative intensity $\Lambda(t)$, the number of events that occurred between time zero and t is a random variable that is Poisson distributed with parameter $\Lambda(t)$. The intensity function $\lambda(t)$ of a NHPP can increase or decrease abruptly and the shift point in time is the change-point. Lawless and Zhan (1998) pointed out that the constant piecewise function is an good alternative when the form of the intensity function is unknown.

Assume that the number of events till time t is a NHPP for each driver with cumulative intensity $\Lambda(t)$. The number of events that occurred between time zero and t is a random variable that is Poisson distributed with parameter $\Lambda(t)$. The intensity function $\lambda(t)$ of a

NHPP can increase or decrease abruptly and the shift point in time is the change-point. Lawless and Zhan (1998) pointed out that the constant piecewise function is an good alternative when the form of the intensity function is unknown. Accordingly, I assume the intensity function $\lambda_j(t)$ of the NHPP is constant piecewise with a single unknown change-point $\tau_j \in (0, C_j)$, $j = 1, 2, \cdots, m$. There drivers fall in K clusters with change-point values $\mu_1 < \mu_2 < \cdots < \mu_K$ and intensity rates before and after the change-points to be $(\lambda_{b1}, \lambda_{a1}), \cdots, (\lambda_{bK}, \lambda_{aK})$. That is, the drivers in the same cluster share identical change-points and intensity rates, considering the relatively smaller number of events for NDS.

3.2.2 Hierarchical Bayesian framework

Let z_j be the latent membership of the j^{th} driver, $z_j \in \{1, 2, \cdots, K\}$. $z_j \sim Generalized\ Bernoulli(\boldsymbol{\pi}_j)$ with the point mass function $f(z_j|\boldsymbol{\pi}_j, K) = \prod_{k=1}^{K} \pi_{jk}^{\delta_k(z_j)} = \sum_{k=1}^{K} \pi_{jk}\delta_k(z_j)$ (Murphy, 2012, p.35), where $\delta_k(z_j)$ is 1 when $z_j = k$ and 0 otherwise. The mixture proportion $\boldsymbol{\pi}_j = (\pi_{j1}, \pi_{j2}, \cdots, \pi_{jK})$, $0 \leq \pi_{jK} \leq 1$, $\sum_{k=1}^{K} \pi_{jk} = 1$.

$\boldsymbol{\pi}_j|\alpha_0, K \sim Dir(\alpha_0/K \cdot \mathbf{1}_K)$, where $Dir(\cdot)$ is a Dirichlet distribution, and $\mathbf{1}_K$ is a vector of ones with length K. The symmetric Dirichlet distribution is chosen because of no prior information on mixture proportions. The Dirichlet distribution $Dirichlet(\alpha_1, \alpha_2, \cdots, \alpha_K)$ is a multivariate extension of a Beta distribution, where each parameter is a positive real number (Kotz et al., 2000, Chap. 49). $K > 1$ is the number of categories. The support of a Dirichlet distribution is an open $(K-1)$-dimensional simplex defined by $\left\{ x_k \in (0,1) \sum_{k=1}^{K} x_k = 1 \right\}$. The Dirichlet distribution is the conjugate distribution of a multinomial distribution. Let $i_{jk} = 1$ if $z_j = k$ and $i_{jk} = 0$ otherwise. $\boldsymbol{i}_j = (i_{j1}, i_{j2}, \cdots, i_{jK}) \sim multinomial(1, \boldsymbol{\pi}_j)$.

I assume $\mu_k \sim G_0(\theta)$, $k = 1, 2, \cdots, K$, where $G_0(\theta)$ is a continuous uniform distribution $Unif(0, \theta)$ and θ is a fixed upper bound. This method assumes that μ_k is independent from λ_{bk} and λ_{ak}, and also that λ_{bk} and λ_{ak} are conditionally independent given μ_k. Considering Gamma priors for λ_{bk} and λ_{ak} given μ_k, $\lambda_{bk}|\mu_k \sim Gamma(a_{1k}, b_{1k})$, $\lambda_{bk}|\mu_k \sim Gamma(a_{2k}, b_{2k})$, where $Gamma(a, b)$ is a Gamma distribution with mean a/b and variance a/b^2. The joint distribution is:

$$f(\lambda_{bk}, \lambda_{ak}, \mu_k) \propto f(\lambda_{bk}, \lambda_{ak}|\mu_k) f(\mu_k) \propto \lambda_{bk}^{a_{1k}-1} e^{-b_{1k}\lambda_{bk}} \lambda_{ak}^{a_{2k}-1} e^{-b_{2k}\lambda_{ak}} f(\mu_k),$$

$\lambda_{bk} > 0,\ \lambda_{ak} > 0.$

The prior of α_0 is $f(\alpha_0)$, which can be a gamma distribution.

The intensity function for the j^{th} driver is:

$$\lambda_j(t) = \lambda_{bz_j}(0 \le t \le \tau_j) + \lambda_{az_j} I(t > \tau_j),\ j = 1, 2, \cdots, m.$$

The likelihood for the j^{th} driver (Thompson, 1988) is

$$L_j(\boldsymbol{\lambda}_j, \tau_j | \boldsymbol{X}_j) = exp[-\Lambda(C_j)] \prod_{i=1}^{n_j} \lambda_j(t_{ji}).$$

Hence the complete likelihood for the j^{th} driver is:

$$\begin{aligned} L_j(\boldsymbol{\lambda_b}, \boldsymbol{\lambda_a}, \boldsymbol{\mu}, \boldsymbol{\pi_j} | z_j, \boldsymbol{X}_j) &= \left\{ exp[-\Lambda(C_j)] \prod_{i=1}^{n_j} \lambda_j(t_{ji}) \right\} f(z_j | \boldsymbol{\pi}_j, K) \\ &= \sum_{k=1}^{K} exp\left[-\lambda_{bk}\mu_k - (C_j - \mu_k)\lambda_{ak}\right] \lambda_{bk}^{N_j^{(b)}} \lambda_{ak}^{N_j^{(a)}} \pi_{jk} \delta_k(z_j) \\ &= \prod_{k=1}^{K} \left\{ exp\left[-\lambda_{bk}\mu_k - (C_j - \mu_k)\lambda_{ak}\right] \lambda_{bk}^{N_j^{(b)}} \lambda_{ak}^{N_j^{(a)}} \pi_{jk} \right\}^{\delta_k(z_j)}, \end{aligned}$$

where $N_j^{(b)}$ is the number of events for the j^{th} driver before change-point τ_j and $N_j^{(a)}$ after, $\boldsymbol{\lambda_b} = (\lambda_{b1}, \cdots, \lambda_{bK})$, and $\boldsymbol{\lambda_a} = (\lambda_{a1}, \cdots, \lambda_{aK})$, $\boldsymbol{\mu} = (\mu_1,\ \mu_2,\ \cdots,\ \mu_K)$ and $\boldsymbol{X}_j$ is the vector of event times for the j^{th} driver.

Let S_k be the index set of j in the k^{th} cluster, and the number of drivers in the k^{th} cluster is m_k. The complete likelihood for the drivers in the k^{th} cluster is:

$$L_{(k)}(\lambda_{bk}, \lambda_{bk}, \mu_k, \boldsymbol{\pi}_{(k)}|\boldsymbol{X}_{(k)})$$
$$= exp\left\{-(\lambda_{bk} - \lambda_{ak})m_k\mu_k - \lambda_{ak}\sum_{j\in S_k} C_j\right\} \lambda_{bk}^{\sum_{j\in S_k} N_j^{(b)}} \lambda_{ak}^{\sum_{j\in S_k} N_j^{(a)}} \prod_{j\in S_k} \pi_{jk}.$$

The posterior distribution is:

$$f(\boldsymbol{\lambda_b}, \boldsymbol{\lambda_a}, \boldsymbol{\mu}, \boldsymbol{\Pi}|\boldsymbol{z}, \boldsymbol{X}) \propto \left\{\prod_{j=1}^{m} L_j(\boldsymbol{\lambda_b}, \boldsymbol{\lambda_a}, \boldsymbol{\mu}, \boldsymbol{\pi_j}|z_j, \boldsymbol{X})\right\} f(\boldsymbol{\lambda_b})f(\boldsymbol{\lambda_a})f(\boldsymbol{\mu})f(\boldsymbol{\Pi}). \tag{3.1}$$

Therefore, the full conditional distribution for $\boldsymbol{\pi}_j$ is:

$$f(\boldsymbol{\pi}_j|\boldsymbol{X}_j) = Dir(\alpha_0/K + i_{j1}, \alpha_0/K + i_{j2}, \cdots, \alpha_0/K + i_{jK}). \tag{3.2}$$

The full conditional distribution for μ_k is:

$$f(\mu_k|\boldsymbol{X}_{(k)}) \propto G_0(\theta)L_{(k)}. \tag{3.3}$$

In the Bayesian Markov chain Monte Carlo (MCMC) algorithm, I sample from the full conditional distribution for z_j:

$$f(z_j|\boldsymbol{\pi}_j, \boldsymbol{\mu}, \boldsymbol{X}_j) \propto \sum_{k=1}^{K} L_j(\tau_j = \mu_k|z_j = k, X_j)\delta_k(z_j). \tag{3.4}$$

Based on Bayes' theorem, the conditional distributions given data of the k^{th} cluster $\boldsymbol{X}_{(k)}$ are as follows:

$$\begin{aligned} f(\lambda_{bk}|\mu_k, \lambda_{ak}, \boldsymbol{X}_{(k)}) \propto f(\lambda_{bk}|\mu_k)L_{(k)} &= \frac{\int_0^\infty f(\lambda_{bk}, \lambda_{ak}, \boldsymbol{\tau}_{(k)})d\lambda_{ak}}{f(\mu_k)} L_{(k)} \\ &\sim Gamma(a_{1k} + \sum_{j\in S_k} N_j^{(1)}, b_{1k} + \sum_{j\in S_k} \tau_j), \end{aligned} \tag{3.5}$$

$$f(\lambda_{ak}|\mu_k, \lambda_{bk}, \boldsymbol{X}_{(k)}) \propto L_{(k)}\lambda_{ak}^{a_{2k}-1}e^{-b_{2k}\lambda_{ak}}$$
$$\sim Gamma(a_{2k} + \sum_{j\in S_k} N_j^{(2)}, b_{2k} + \sum_{j\in S_k}(C_j - \tau_j)). \tag{3.6}$$

Let $\boldsymbol{\Pi} = (\boldsymbol{\pi}_1,\ \boldsymbol{\pi}_2,\ \cdots,\ \boldsymbol{\pi}_m)$. The full conditional for α_0 is

$$f(\alpha_0|\boldsymbol{\Pi}, K) \propto f(\alpha_0)f(\boldsymbol{\Pi}|\alpha_0, K) \propto f(\alpha_0)\prod_{j=1}^{m} f(\boldsymbol{\pi}_j|\alpha_0, K). \tag{3.7}$$

3.2.3 Bayesian sampling from the posterior

The collapsed Gibbs sampler (Liu, 2003) combined with rejection sampling (Neal, 2003) is used to sample from the full conditional distributions that have no closed form. The initial values for the MCMC are as follows: $\boldsymbol{\mu}^{(0)}$ and $\alpha_0^{(0)}$ are randomly generated from their priors; the initial values of λ_{bk} and λ_{ak} are the average incident rates of each cluster before and after the change-points; the subjects are randomly assigned to the clusters. For $j = 1,\ 2,\ \cdots,\ m,\ k = 1,\ 2,\ \cdots,\ K$:

Algorithm 1.

For $t = 1,\ 2,\ \cdots,\ B_t$ (B_t is the number of iterations),

1. For $j = 1,\ 2,\ \cdots,\ m$,
   generate $\boldsymbol{\pi}_j^{(t)}$ from $f(\boldsymbol{\pi}_j|\boldsymbol{X})$ in Eq. 3.2,

$$\boldsymbol{\pi}_j^{(t)} \sim Dir(\alpha_0^{(t-1)}/K + i_{j1}^{(t-1)}, \alpha_0^{(t-1)}/K + i_{j2}^{(t-1)}, \cdots, \alpha_0^{(t-1)}/K + i_{jK}^{(t-1)}).$$

2. For $k = 1,\ 2,\ \cdots,\ K$,
   sample $\mu_k^{(t)}$ from $\mu_k|\boldsymbol{X} \propto G_0(\theta)L_{(k)}(\lambda_{bk}^{(t-1)}, \lambda_{ak}^{(t-1)}, \boldsymbol{\tau}_{(k)} = \mu_k \cdot \mathbf{1}|\boldsymbol{z} = k \cdot \mathbf{1})$, using rejection sampling. $\boldsymbol{\tau}_{(k)}$ is the vector including all the change-points with value μ_k.

3. For $j = 1,\ 2,\ \cdots,\ m$,
   generate $z_j^{(t)}$ from $f(z_j|\boldsymbol{\pi}_j^{(t)}, \boldsymbol{\mu}^{(t)}, \boldsymbol{X})$ in Eq. 3.4.

4. For $j = 1,\ 2,\ \cdots,\ m$,
   update $\tau_j^{(t)}$ and $\boldsymbol{i}_j^{(t)}$ based on $z_j^{(t)}$ and $\boldsymbol{\mu}^{(t)}$.

5. For $k = 1,\ 2,\ \cdots,\ K$,
   generate $\lambda_{bk}^{(t)}$ from $f(\lambda_{bk}|\lambda_{ak}^{(t-1)}, \boldsymbol{\tau}_{(k)}^{(t)}, \boldsymbol{X})$ in Eq. 3.5; generate $\lambda_{ak}^{(t)}$ from $f(\lambda_{ak}|\lambda_{bk}^{(t)}, \tau_{(k)}^{(t)}, \boldsymbol{X})$ in Eq. 3.6.

6. Generate $\alpha_0^{(t)}$ from $f(\alpha_0|\mathbf{\Pi})$ in Eq. 3.7 using rejection sampling.

Five chains with widely dispersed initial values are run until the Gelman-Rubin statistics are less than 1.2 for all scalar summaries (Gelman and Rubin, 1992). The DICs (Spiegelhalter et al., 2002) of the models with different number of clusters are compared to determine the number of subgroups. Each subject is assigned to the cluster with the maximum class membership probability.

3.3 Simulation Studies

A simulation study is conducted to examine the model performance under different configurations.

Data generation from a NHPP with piecewise constant intensity functions is based on the inter-event times' distribution (Klein and Roberts, 1984). Given the previous event times $Y_0 = y_0 = 0,\ Y_1 = y_1,\ \cdots,\ Y_{i-1} = y_{i-1}$, the i^{th} inter-event time $X_i = Y_i - Y_{(i-1)}$ for each subject has the cumulative density function (CDF):

$$\begin{aligned} F_i(x) &= Pr\left[X_i \leq x | Y_p = y_p,\ p = 1, 2, \cdots, i-1\right] \\ &= 1 - exp\left[\mathit{\Lambda}(y_{i-1}) - \mathit{\Lambda}(y_{i-1} + x)\right]. \end{aligned} \tag{3.8}$$

The event time y_{i+1} is the summation of the $(i+1)^{th}$ inter-event time with CDF F_i and the i^{th} event time. The algorithm to generate data from a NHPP is as follows:

1. $i = 0$, set $y_{(i)} = 0$;

2. Sample from F_i in Eq. (3.8) to obtain $x_{(i+1)}$;

3. $y_{(i+1)} = y_{(i)} + x_{(i+1)}$;

4. $i = i + 1$, return to step (2).

Repeat the above process until y_{i+1} is larger than the censoring time for each individual. The event times are $y_1, \ y_2, \ \cdots, \ y_i$ for one subject.

3.3.1 Simulation configurations

Seven settings with different change-points, intensity rates, mixture proportions, sample sizes and number of clusters are considered. Specifically, I compare the results for the first setting with other settings to check how different senorios would affect the results. The total number of drivers m for each data set is 40 except for the forth setting. In the first setting, each driver has the same probability to be from different clusters: $\pi_{j1} = \pi_{j2} = \pi_{j3} = 1/3$. The second setting is the same as the first one except that the first two clusters have identical intensity rates: $\lambda_{b1} = \lambda_{b2}, \ \lambda_{a1} = \lambda_{a2}$. The third setting is the same as the first one except that the first two clusters have identical change-points: $\mu_1 = \mu_2$. The forth setting is the same as the first one except that there are 80 subjects: $m = 80$. The fifth setting is the same as the first one except that the mixture proportions are not equal: $(\pi_{j1}, \pi_{j2}, \pi_{j3}) = (10\%, 45\%, 45\%)$. In the sixth setting, the parameter values are close to the estimates of NTDS application in Section 3.4, and $\pi_{j1} = \pi_{j2} = \pi_{j3} = 1/3$. The seventh setting has only one cluster: $K = 1$. The censoring times are uniform between 450 and 500 hours.

The priors and initial values for the MCMC are as follows: the prior of the change-points $G_0(\theta)$ is a continuous uniform distribution $Unif(0, 400)$; $1/\lambda_{bk}$ and $1/\lambda_{ak}$ are the priors for the intensity rates λ_{bk} and λ_{ak}; α_0 is 2.2, which is obtained from screening simulation; the initial values of the change-point μ_k is from $Unif(0, 400)$; the subjects are randomly assigned to the clusters; the initial values of λ_{bk} and λ_{ak} are the average incident rates of each cluster before and after the change-points.

Under each parameter setting, $B = 200$ data sets are generated. The method in Section 2.2 with the number of clusters from one to seven is applied to each data set. I run 15,000 iterations for each MCMC chain, discard the first 5,000 iterations, and thin the sample after burn-in by using every fifth observation. The percentages of the data sets with correctly identified number of clusters and the parameter estimates given correct number of clusters under each setting are recorded. I average the estimates of the B data sets, compute the Root mean square error (RMSE) for a parameter τ by $\sqrt{(1/B)\sum_{k=1}^{B}(\widehat{\tau}-\tau)^2}$, and |bias (%)| by $(1/B)\sum_{k=1}^{B}|\widehat{\tau}-\tau|/\tau \times 100\%$, estimate the coverage probability by the 95% highest posterior density (HPD) interval, and average the percentage of correctly grouped subjects.

3.3.2 Simulation results

The percentages of the data sets with correctly estimated number of clusters (%) are 95.0, 92.5, 90.5, 95.5, 93.0, 87.0 and 97.0, and the average percentage of correctly grouped subjects (%) are 96.23, 97.45, 98.51, 97.21, 94.35, 80.28, 100.00 correspondingly to the seven settings.

Tables 3.1 and 3.2 show the estimation results for the seven settings. For the first setting where the parameters are widely dispersed, the RMSE is reasonable, the absolute percentage bias (%) is within 0.1%, and the coverage probabilities are close to 95.0 %. For the second and third settings where two clusters share the same change-point value or intensity rates, the estimations are accurate, however, the percentages of the data sets with correctly estimated number of clusters are smaller compared with the first setting. For the forth setting when the sample size m is larger than the first setting, the RMSE is much smaller than the first setting. For the fifth setting, both the RMSE and |bias (%)| are larger for the first cluster, which has the smallest mixture proportion. For the sixth setting, the change-points and intensity rates are less dispersed and smaller, the |bias (%)| are much larger and the coverage probability are lower than that of previous settings. The results are satisfactory for the last setting with one cluster.

Table 3.1: Simulation results for the first three settings

Setting	Parameter	True value	Average of estimates	RMSE	|Bias (%)|	Coverage probability (%)
	μ_1	110	110.31	6.82	0.28	96.5
	μ_2	220	220.76	9.46	0.34	97.0
	μ_3	330	329.74	14.06	0.08	95.0
	λ_{b1}	100	100.93	17.26	0.93	91.5
1	λ_{a1}	200	198.91	7.19	0.54	93.5
	λ_{b2}	250	248.22	16.54	0.71	91.0
	λ_{a2}	150	151.36	10.92	0.91	93.0
	λ_{b3}	300	300.48	9.86	0.16	92.5
	λ_{a3}	200	198.47	11.51	0.76	94.0
	μ_1	110	114.31	17.90	3.92	92.5
	μ_2	220	220.90	11.98	0.41	95.0
	μ_3	330	328.92	10.80	0.33	95.5
	λ_{b1}	250	249.18	15.44	0.33	93.5
2	λ_{a1}	150	150.45	7.52	0.30	92.0
	λ_{b2}	250	249.67	12.79	0.13	94.5
	λ_{a2}	150	149.67	9.91	0.22	90.5
	λ_{b3}	300	299.24	9.67	0.25	93.5
	λ_{a3}	200	199.37	11.59	0.31	95.0
	μ_1	110	107.68	6.28	2.11	92.5
	μ_2	110	112.48	9.55	2.25	92.5
	μ_3	330	329.84	9.48	0.05	95.0
	λ_{b1}	100	99.73	6.03	0.27	96.5
3	λ_{a1}	200	199.08	5.05	0.46	92.5
	λ_{b2}	250	247.39	10.42	1.04	94.0
	λ_{a2}	150	149.81	4.90	0.13	90.5
	λ_{b3}	300	300.61	5.84	0.20	96.0
	λ_{a3}	200	199.92	7.04	0.04	97.0

Table 3.2: Simulation results for other four settings

Setting	Parameter	True value	Average of estimates	RMSE	|Bias (%)|	Coverage probability (%)
	μ_1	110	109.98	3.20	0.02	95.5
	μ_2	220	220.15	4.41	0.07	93.0
	μ_3	330	329.57	5.13	0.13	93.0
	λ_{b1}	100	99.96	6.27	0.04	94.0
4	λ_{a1}	200	199.68	4.58	0.16	94.5
	λ_{b2}	250	250.12	7.13	0.05	94.0
	λ_{a2}	150	150.35	5.09	0.23	97.0
	λ_{b3}	300	299.92	6.40	0.03	94.5
	λ_{a3}	200	200.35	8.24	0.18	93.0
	μ_1	110	108.50	16.74	1.37	96.5
	μ_2	220	215.75	5.76	1.93	93.0
	μ_3	330	328.36	13.79	0.50	95.5
	λ_{b1}	100	111.28	27.62	11.28	90.5
5	λ_{a1}	200	199.17	13.04	0.41	93.5
	λ_{b2}	250	254.43	10.13	1.77	93.0
	λ_{a2}	150	150.88	12.00	0.59	91.5
	λ_{b3}	300	299.80	8.96	0.07	93.0
	λ_{a3}	200	198.39	10.12	0.81	95.5
	μ_1	50	59.52	13.84	19.05	89.0
	μ_2	110	108.27	11.46	1.58	89.0
	μ_3	150	157.51	14.62	5.01	89.0
	λ_{b1}	48	41.02	15.62	14.54	76.5
6	λ_{a1}	54	42.67	21.59	20.99	70.0
	λ_{b2}	27	28.81	10.02	6.71	86.5
	λ_{a2}	13	18.71	15.4	43.89	85.0
	λ_{b3}	20	26.98	13.38	34.91	86.5
	λ_{a3}	11	17.62	16.31	60.19	87.0
	μ	220	219.68	2.99	0.15	92.5
7	λ_b	300	299.52	5.13	0.16	95.5
	λ_a	200	200.20	4.57	0.10	93.5

Overall, when the difference in change-points or intensity rates among clusters are larger, or when the change-points and intensity rates are more dispersed and larger, the estimates will be more accurate. When the mixture proportion of the clusters is smaller, the estimate will be less stable. Larger sample size results in smaller variation in the estimators. The model gives fine results under different scenarios, and is not sensitive to priors and initial values.

3.4 Application to the NTDS

The model is applied to the NTDS data. The 38 teenagers with CNC events are considered.

3.4.1 NTDS results

The prior of μ's is a continuous uniform distribution $Unif(0, 263)$. I choose 263 as the upper bound of the change-points, because the average of the censoring time for the 38 teenagers is 263.07 hours. $\lambda_{bk} \sim Gamma(30, 1)$, $\lambda_{ak} \sim Gamma(10, 1)$, $k = 1,\ 2, \cdots,\ K$. $\alpha_0 \sim Gamma(2.6, 1)$ based on preliminary NTDS application results.

The initial values of μ_2 till μ_K are sampled uniformly from 0 to 263. $\mu_1^{(0)}$ is set to be smaller than the minimum of the censoring times, because the change-point of each subject must be smaller than his or her censoring time to guarantee the MCMC convergence. $\lambda_{bk}^{(0)}$, $\lambda_{ak}^{(0)}$ and $\mu_1^{(0)}$ are 30, 10 and 65 respectively, which are close to the estimated intensity rates and change-point of the identical-intensity model with one change-point in Li et al. (2015). Each subject is assigned randomly to the clusters. The initial value of α_0 is 2.6.

For each fixed number of clusters, I generate a MCMC chain of 70,000 iterations, discard the first 10,000 iterations, and use every sixth observation after burn-in. The DICs when the number of clusters is from one to seven are calculated.

The model with three clusters is chosen for NTDS data. The DIC is 2493.6 when the number of clusters is one, 2464.3 for two clusters and 2378.9 for three clusters. When the number of clusters are larger than three, though the DIC can be smaller, the differences between the intensity estimates before change-point and after can be smaller than one CNC events per teenager per 1,000 hours.

Table 3.3 shows the posterior means, SDs, and 95 % HPD intervals for the model with three clusters. There are 14, 13 and 11 teenagers in the three clusters correspondly. For the first cluster, the change-point is 52.30 hours and the intensity before the change-point is 48.94 CNC events per teenager per 1,000 hours and 54.22 after the change-point. For the second cluster, the change-point is 108.99 hours and the intensity before the change-point is 27.89 and 13.05 after. For the third cluster, the change-point is 150.20 hours and the intensity before the change-point is 18.76 and 11.98 after.

Table 3.3: Posterior estimates of the model with three clusters

Parameters	Posterior mean	Posterior SD	95% HPD interval
μ_1	52.30	12.12	(29.38, 67.80)
μ_2	108.99	14.04	(59.06, 113.70)
μ_3	150.20	17.23	(114.24, 177.53)
λ_{b1}	48.94	10.93	(28.37, 67.99)
λ_{a1}	54.22	5.60	(44.35, 64.37)
λ_{b2}	27.89	9.86	(9.72, 44.42)
λ_{a2}	13.05	3.68	(6.56, 19.55)
λ_{b3}	18.76	7.05	(7.17, 31.47)
λ_{a3}	11.98	4.10	(5.18, 19.56)
α_0	2.58	1.25	(0.59, 4.94)

For the model with three clusters, the ratio of female vs. male are close to one in each cluster. Fig. 3.1 shows the cumulative event plot vs. the estimated cumulative intensity functions of the three clusters. The classification is reasonable. Fig. 3.2 shows the posterior distributions of the parameters with kernel fitting. The posterior distributions are all uni-model, and most of them are right skewed. Fig. 3.3 shows the histogram of the mixture proportion estimates of all the drivers with kernel fitting. For most of the drivers, the maximum mixture proportion is domeninat among the three mixture proportions.

Overall, for the intensity rates before and after the change-point, it increases for the

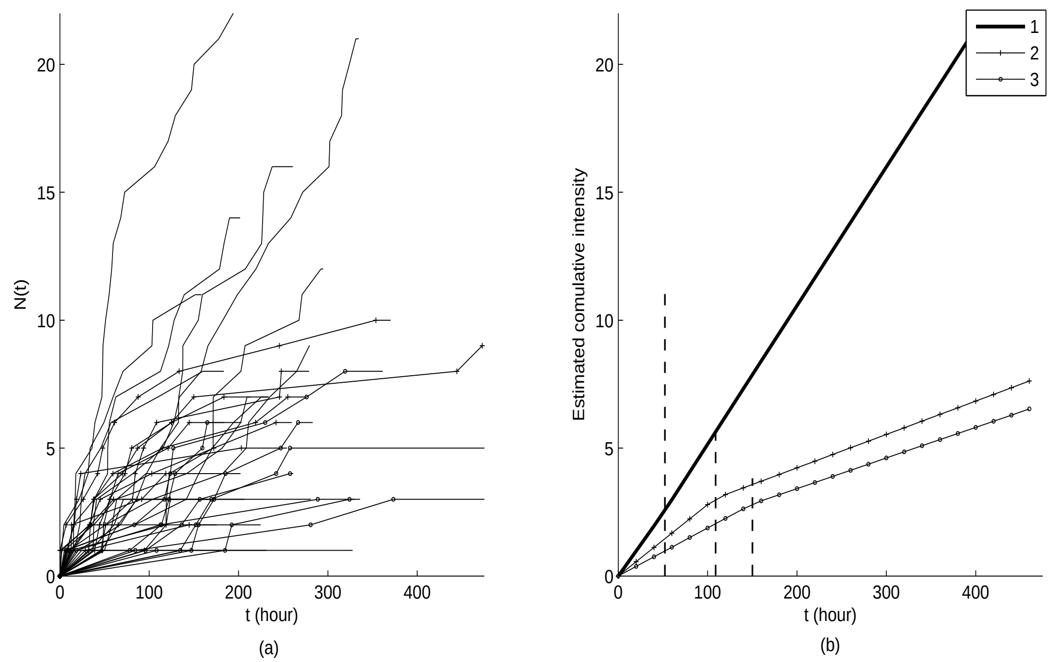

Figure 3.1: The model with three clusters: (a) The cumulative event plot of NTDS. The symbols mark the event-times. (b) The estimated cumulative intensity functions of the three clusters. The vertical dashed lines show the estimated change-points, and the heights represent the average number of events in each cluster.

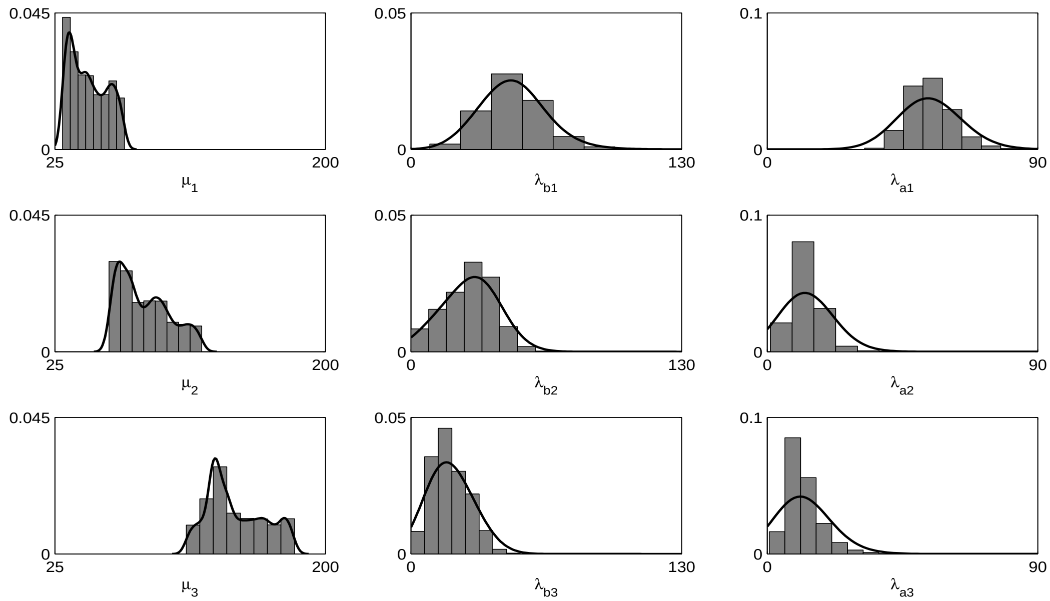

Figure 3.2: Posterior distributions of the parameters with kernal fitting

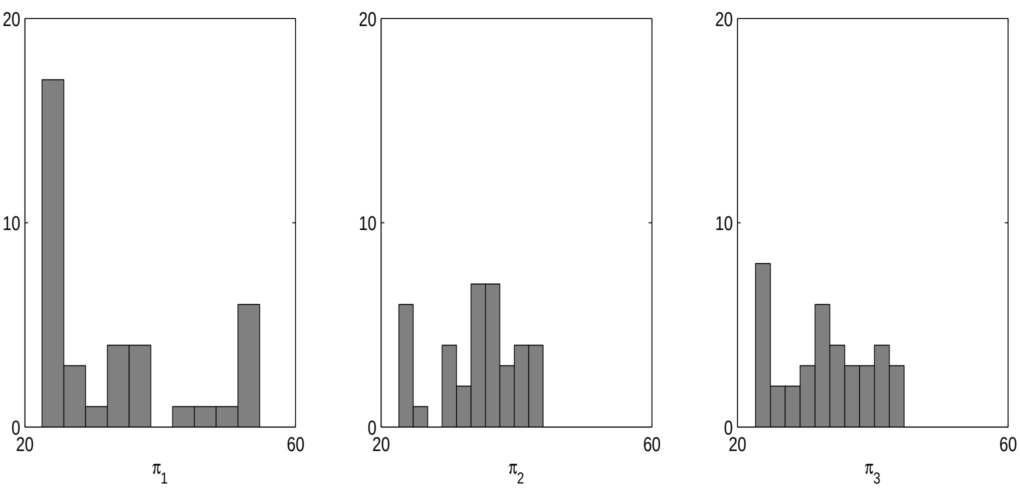

Figure 3.3: Histogram of the mixture proportions

first cluster while decreases for the other clusters. The change-point of the first cluster is the smallest and the average intensity rate is the highest. Comparing with the average cumulative driving time per month per teenager in Table 1.1 , the changes happened during the fourth, eighth, and eleventh month accordingly.

3.4.2 Comparison with previous work

I compare the results in this chapter with Guo et al. (2013). Guo et al. (2013) grouped the teenagers of NTDS into three clusters based on overall CNC rates by employing a K-means clustering method, and estimated the CNC rates over time by mixed effect Poisson models. This chapter clusters the teenagers and detect the change in one model. Table 3.4 shows the contingency table of the NTDS clustering results in Guo et al. (2013) and in this research. The teenagers in the first cluster mainly fall in the high-risk group, and the teenagers in the second cluster mainly fall in the moderate-risk group, while the teenagers in the third cluster mainly fall in the low-risk group. Twenty eight teenagers out of 38 are grouped similarly by the two methods. The average intensity rate for the first cluster by BFMM is the highest among the three clusters and the change happened at an earlier stage. It is reasonable that this cluster is corresponding to the high-risk group, and vice versa. Guo et al. (2013) showed the change-points by three-month intervals, while the method in the chapter evaluates the change-points in actual driving time. The change-points are consistent between the two methods.

Table 3.4: Comparison between two methods

BFMM \ K-means	Low-risk	Moderate-risk	High-risk	Total
Cluster 1	1	3	10	14
2	0	10	3	13
3	8	3	0	11
Total	9	6	13	38

3.5 Conclusion and Discussion

The driving risk of novice teenage drivers decreased significantly in the early period after first licensure. How much experience is required before relatively safe driving can be quantified by the change-points in crash rates.

Previous studies suggested that the change-point was about six months of driving after first licensure. However, teenagers have different driving time during the same period. As a result, identifying the risk change-point in driving hours is important.

Subgroups might exist among the teenagers with different patterns of risk change. Classifying the teenagers based on the risk pattern accommodates the heterogeneity among drivers and provides more insight in teenagers' driving behaviour.

This chapter is based upon a hierarchical Bayesian finite mixture model. According to the simulation studies, the model is robust to different parameter settings and give decent results under different scenarios.

The model is applied to the NTDS data. After the model selection, the change-point model with three clusters is chosen. The intensity rates increase for the first cluster and decrease for the other clusters after the change-points. The results are consistent with Guo et al. (2013), while the method in this research has several advantages.

In this research, model selection is conducted to determine the best number of clusters. Teh (2010) pointed out that an alternative was the Bayesian non-parametric approach. The non-parametric method to automatically detect the number of clusters with a Dirichlet process prior (Antoniak, 1974) will be considered in the future. Furthermore, I would like to incorporate covariates and predict the clustering.

The methodology has many applications across disciplines including the study of medical treatment effects, the hazard pattern of certain diseases, stock market analysis and quality control.

Chapter 4 A NON-PARAMETRIC BAYESIAN CHANGE-POINTS MODEL FOR DETECTING DRIVING RISK CHANGES

4.1 Introduction

The substantial variation in driving risk has been documented in many studies (Rolls et al., 1991; Ulleberg, 2001; Dingus et al., 2006; Musselwhite, 2006; Klauer et al., 2009). It is reasonable to expect clustering considering the heterogeneity among drivers. A latent class analysis was conducted in Chapter 3 to address the underlying unknown structure. Traditional latent class modelling contains model selection to choose the best number of clusters. Neal (2000) and Teh (2010) pointed out that an alternative was the Bayesian non-parametric approach.

For the Bayesian non-parametric inference, the prior and posterior distributions are stochastic processes instead of a prior from a parametric family. The representations grew with unbounded complexity when observing more data (Walker et al., 1999; Hjort et al., 2010).

In Bayesian non-parametric models, Dirichlet process (DP) is frequently used, especially in clustering problems when the number of clusters is unknown a priori (Antoniak, 1974). This type of model is the Dirichlet process mixture model (DPMM) (Ferguson, 1983). Because clusters exist among the teenagers while the number of clusters is unknown, the Bayesian non-parametric model with a DP prior would be promising.

A large part of the literatures on DPMM used MCMC to sample from the posterior distribution. Neal (1992), Escobar (1994), and Bush and MacEachern (1996) developed Gibbs sampling methods for DPMM with conjugate priors. West et al. (1994), Escobar and West (1995), MacEachern and Müller (1998), and Walker and Damien (1998) devised MCMC approaches when using non-conjugate priors. Neal (2000) reviews MCMC methods in DPMM and extended Gibbs sampling to sample from the posteriors.

Chapter 4 proposes a DPMM allowing change-points to vary among drivers. A MCMC algorithm is developed to sample from the posterior distributions. Then the K-means clustering is used to group the subjects. Each subject has one unknown change-point, and all the subjects share the identical intensity rates. These change-points fall in several clusters while the number of clusters, the cluster parameters and the cluster assignments are unknown. The model is applied to both simulated data and the NTDS data.

The rest of the chapter is organized as follows. Section 4.2 is the methodology for detecting the change-points by DPMM. Simulation study is in Section 4.3 and NTDS data analysis is in Section 4.4. Section 4.5 is the conclusion and discussion.

4.2 The Bayesian Change-Point Model with a Dirichlet Process Prior

Clusters exist among the teenagers while the number of clusters is unknown. Assume that the event-counting process is a Poisson process and the intensity is constant piecewise with a single change-point. Automatic clustering is expected in the Dirichlet process mixture model.

4.2.1 Methodology

There is at least one CNC event per teenager and the total number of events for the j^{th} driver is n_j. The total number of events for all drivers during the study period is $N = \sum_{j=1}^{m} n_j$. The total driving time for the j^{th} subject is C_j during the study. Assume that $N_j(t)$ (the number of events till time t) is a Poisson process for $j = 1, \cdots, m$, m is the total number of drivers. The intensity function of the Poisson process is constant piecewise.

The model proposed in this chapter assumes one change-point $\tau_j \in (0, C_j)$ for each driver and the change-point varies among drivers, $j = 1, 2, \cdots, m$, $\boldsymbol{\tau} = (\tau_1, \tau_2, \cdots, \tau_m)$. C_j is the censoring time for the j^{th} driver. When all the drivers share the same intensity rates λ_b and

λ_a, the intensity function for the j^{th} driver is:

$$\lambda_j(t) = \lambda_b I(0 \leq t \leq \tau_j) + \lambda_a I(t > \tau_j),\ j = 1, 2, \cdots, m.$$

τ_j follows a general distribution G while G is assigned a DP prior:

$$\tau_j | G \sim G, G \sim DP(\alpha_0, G_0(\theta)).$$

The probability density function of α_0 and θ are $f(\alpha_0)$ and $f(\theta)$ while α_0 and θ are independent.

4.2.2 Sampling scheme

The likelihood of all drivers is:

$$L(\lambda_b, \lambda_a, \boldsymbol{\tau} | \boldsymbol{X}) = exp\left\{ -(\lambda_b - \lambda_a) \sum_{j=1}^{m} \tau_j - \lambda_a \sum_{j=1}^{m} C_j \right\} \lambda_b^{\sum_{j=1}^{m} N_j^{(1)}} \lambda_a^{\sum_{j=1}^{m} N_j^{(2)}}, \tag{4.1}$$

where $N_j^{(1)}$ is the number of events for the j^{th} driver before change-point τ_j and $N_j^{(2)}$ after. $\boldsymbol{X}$ is the data. The posterior distribution is:

$$f(\lambda_b, \lambda_a, \boldsymbol{\tau} | \boldsymbol{X}) \propto L(\lambda_b, \lambda_a, \boldsymbol{\tau} | \boldsymbol{X}) f(\lambda_b) f(\lambda_a) f(\boldsymbol{\tau}) \tag{4.2}$$

Assume that the change-point vector $\boldsymbol{\tau}$ is independent from λ_b and λ_a while λ_b is conditionally independent from λ_a given $\boldsymbol{\tau}$. Consider Gamma priors for λ_b and λ_a given $\boldsymbol{\tau}$, $\lambda_b | \boldsymbol{\tau} \sim Gamma(a_1, b_1)$, $\lambda_a | \boldsymbol{\tau} \sim Gamma(a_2, b_2)$, where $Gamma(a, b)$ is a Gamma distribution with mean a/b and variance a/b^2.

$$f(\lambda_b, \lambda_a, \boldsymbol{\tau} | \alpha_0, \theta) \propto f(\lambda_b, \lambda_a | \boldsymbol{\tau}) f(\boldsymbol{\tau} | \alpha_0, \theta) \propto \lambda_b^{a_1 - 1} e^{-b_1 \lambda_b} \lambda_a^{a_2 - 1} e^{-b_2 \lambda_a} f(\boldsymbol{\tau} | \alpha_0, \theta), \tag{4.3}$$

where $\lambda_b, \lambda_a > 0$. Based on Bayes' theorem, the conditional distributions are as follows:

$$\begin{aligned} f(\lambda_b|\boldsymbol{\tau}, \lambda_a, \boldsymbol{X}) \propto f(\lambda_b|\boldsymbol{\tau})L(\cdot) &= \frac{\int_0^\infty f(\lambda_b, \lambda_a, \boldsymbol{\tau}) d\lambda_a}{f(\boldsymbol{\tau}|\alpha_0, \theta)} L(\cdot) \\ &\propto L(\cdot)\lambda_b^{a_1-1} e^{-b_1\lambda_b} \sim Gamma(a_1 + \sum_{j=1}^{m} N_j^{(1)}, b_1 + \sum_{j=1}^{m} \tau_j), \end{aligned} \tag{4.4}$$

$$f(\lambda_a|\boldsymbol{\tau}, \lambda_b, \boldsymbol{X}) \propto L(\cdot)\lambda_a^{a_2-1} e^{-b_2\lambda_a} \sim Gamma(a_2 + \sum_{j=1}^{m} N_j^{(2)}, b_2 + \sum_{j=1}^{m} (C_j - \tau_j)), \tag{4.5}$$

The joint posterior is:

$$f(\lambda_b, \lambda_a, \boldsymbol{\tau}|\boldsymbol{X}, \alpha_0, \theta) \propto L(\cdot)\lambda_b^{a_1-1} e^{-b_1\lambda_b} \lambda_a^{a_2-1} e^{-b_2\lambda_a} f(\boldsymbol{\tau}|\alpha_0, \theta). \tag{4.6}$$

According to Escobar (1994), the joint distribution of $\boldsymbol{\tau}$ is:

$$f(\tau_1, \tau_2, \cdots, \tau_m|\alpha_0, \theta) = \prod_{j=1}^{m} \frac{\alpha_0 G_0(\tau_j|\theta) + \sum_{p=1}^{j-1} \delta(\tau_p, \tau_j)}{\alpha_0 + j - 1}, \tag{4.7}$$

$\boldsymbol{\tau}_{-j}$ is the vector of change-points for all drivers exclude τ_j. Neal (1998) showed that the distribution of τ_j conditioned on $\boldsymbol{\tau}_{-j}, \alpha_0, \theta$ is:

$$\tau_j|\boldsymbol{\tau}_{-j}, \alpha_0, \theta \sim \frac{\alpha_0 G_0(\tau_j|\theta) + \sum_{p \neq j} \delta(\tau_j, \tau_p)}{\alpha_0 + m - 1}. \tag{4.8}$$

The marginal posterior of $\boldsymbol{\tau}$ is:

$$\begin{aligned} f(\boldsymbol{\tau}|\boldsymbol{X}, \alpha_0, \theta) &= \int_0^\infty \int_0^\infty f(\lambda_b, \lambda_a, \boldsymbol{\tau}|\boldsymbol{X}) d\lambda_b d\lambda_a \\ &\propto f(\boldsymbol{\tau}|\alpha_0, \theta) \frac{\Gamma(a_1 + \sum_{j=1}^{m} N_j^{(1)})\Gamma(a_2 + \sum_{j=1}^{m} N_j^{(2)})}{(b_1 + \sum_{j=1}^{m} \tau_j)^{a_1 + \sum_{j=1}^{m} N_j^{(1)}} (b_2 + \sum_{j=1}^{m} C_j - \sum_{j=1}^{m} \tau_j)^{a_2 + \sum_{j=1}^{m} N_j^{(2)}}}. \end{aligned} \tag{4.9}$$

$\Gamma(\cdot)$ is a Gamma function. The posterior conditional distribution of τ_j is:

$$f(\tau_j|\boldsymbol{\tau}_{-j},\alpha_0,\theta,\boldsymbol{X}) = \frac{f(\boldsymbol{\tau}|\alpha_0,\theta,\boldsymbol{X})}{f(\boldsymbol{\tau}_{-j}|\alpha_0,\theta,\boldsymbol{X}_{-j})} \propto f(\tau_j|\boldsymbol{\tau}_{-j},\alpha_0,\theta)A(\boldsymbol{\tau},\boldsymbol{X}), \tag{4.10}$$

where $A(\boldsymbol{\tau},\boldsymbol{X}) = \dfrac{\Gamma(a_1+\sum\limits_{p=1}^{m}N_p^{(1)})\Gamma(a_2+\sum\limits_{p=1}^{m}N_p^{(2)})}{(b_1+\sum\limits_{p=1}^{m}\tau_p)^{a_1+\sum\limits_{p=1}^{m}N_p^{(1)}}(b_2+\sum\limits_{p=1}^{m}C_p-\sum\limits_{p=1}^{m}\tau_p)^{a_2+\sum\limits_{p=1}^{m}N_p^{(2)}}}$, $\boldsymbol{X}_{-j}$ is the data without the j^{th} driver.

Plug $f(\tau_j|\boldsymbol{\tau}_{-j},\alpha_0,\theta)$ in Eq. 4.8 into the above equation,

$$f(\tau_j|\boldsymbol{\tau}_{-j},\alpha_0,\theta,\boldsymbol{X}) = b\alpha_0 q_0 H(\tau_j|\boldsymbol{\tau}_{-j},\boldsymbol{X}) + b\sum_{p\neq j}A(\boldsymbol{\tau},\boldsymbol{X})\delta(\tau_j-\tau_p), \tag{4.11}$$

$$\text{where } b = [\alpha_0 q_0 + \sum_{p\neq j}A(\boldsymbol{\tau},\boldsymbol{X})]^{-1}, \tag{4.12}$$

$$q_0 = \int G_0(\tau_j|\theta)A(\boldsymbol{\tau},\boldsymbol{X})d\tau_j, \tag{4.13}$$

$$H(\tau_j|\boldsymbol{\tau}_{-j},\theta,\boldsymbol{X}) = G_0(\tau_j|\theta)A(\boldsymbol{\tau},\boldsymbol{X})/q_0. \tag{4.14}$$

q_0 can be computed using numeric method.

$$f(\alpha_0|\boldsymbol{\tau},\theta) = \frac{f(\alpha_0)P(\boldsymbol{\tau}|\alpha_0,\theta)}{\int f(\alpha_0)P(\boldsymbol{\tau}|\alpha_0,\theta)d\alpha_0}. \tag{4.15}$$

$$f(\theta|\boldsymbol{\tau},\alpha_0) = \frac{f(\theta)P(\boldsymbol{\tau}|\alpha_0,\theta)}{\int f(\theta)P(\boldsymbol{\tau}|\alpha_0,\theta)d\theta}. \tag{4.16}$$

In order to sample from the posterior conditional distributions, starting with initial values $\boldsymbol{\tau}^{(0)}$, $\lambda_b^{(0)}$, $\lambda_a^{(0)}$,$\alpha_0^{(0)}$ and $\theta^{(0)}$:

Algorithm 1.

For $t = 1,\cdots,B_t$,

1. For $j = 1,\cdots,m$,

Generate $\tau_j^{(t)}$ from $f(\tau_j|\tau_1^{(t)},\cdots,\tau_{j-1}^{(t)},\tau_{j+1}^{(t-1)},\cdots,\tau_m^{(t-1)},\alpha_0^{(t-1)},\theta^{(t-1)},\boldsymbol{X})$ in Eq. 4.11,

$$\tau_j|\tau_1^{(t)},\cdots,\tau_{j-1}^{(t)},\tau_{j+1}^{(t-1)},\cdots,\tau_n^{(t-1)},\alpha_0^{(t-1)},\theta^{(t-1)},\boldsymbol{X}\begin{cases} =\tau_p, \text{with probability } bA(\boldsymbol{\tau},\boldsymbol{X}), \\ \sim H(\tau_j|\boldsymbol{\tau}_{-j},\boldsymbol{X}), \text{with probability } b\alpha_0 q_0. \end{cases}$$

If $\tau_j \notin [0, C_j]$, resample to generate $\tau_j^{(t)}$.

Rejection sampling can be used to sample from $H(\tau|\boldsymbol{\tau}_{-j},\theta,\boldsymbol{X})$ (Neal, 2003):

(a) Choose a density g (Uniform distribution is chosen);

(b) Find a constant c such that $H(\tau|\boldsymbol{\tau}_{-j},\theta,\boldsymbol{X})/g(\tau) \le c$ for all τ ($c = max\{H/g\}$ is chosen);

(c) Generate a uniform random number u;

(d) Generate a random number v from g;

(e) If $cu \le H(v|\boldsymbol{\tau}_{-j},\theta,\boldsymbol{X})/g(v)$, accept and return v;

(f) Otherwise, reject v and go to step (c);

2. Generate $\lambda_b^{(t)}$ from $f(\lambda_b|\lambda_a^{(t-1)},\boldsymbol{\tau}^{(t)},\boldsymbol{X})$ in Eq. 4.4.

3. Generate $\lambda_a^{(t)}$ from $f(\lambda_a|\lambda_b^{(t)},\tau^{(t)},\boldsymbol{X})$ in Eq. 4.5.

4. Generate $\alpha_0^{(t)}$ from $f(\alpha_0|\boldsymbol{\tau}^{(t)},\theta^{(t-1)})$ in Eq. 4.15.

5. Generate $\theta^{(t)}$ from $f(\theta|\boldsymbol{\tau}^{(t)},\alpha_0^{(t)})$ in Eq. 4.16.

4.2.3 Initial values, hyper parameters and inference

$\boldsymbol{\tau}^{(0)}$ will be set to the individual change-point by the Bayesian finite mixture model in Chapter 3. $\lambda_b^{(0)}$ and $\lambda_a^{(0)}$ will be the average incident rates of all drivers before and after the change-points. $\lambda_b \sim Gamma(\lambda_b^{(0)}, 1)$ and $\lambda_a \sim Gamma(\lambda_a^{(0)}, 1)$. α_0 will be assigned a Gamma prior. The change-points are positive and within the censoring time, $G_0(\theta) = Uniform[0, min(C_j)]$

is chosen. $Uniform[a, b]$ is the uniform distribution on $[a, b]$ and $min(\cdot)$ is the minimum of a vector. After the MCMC sampling, posterior means are the estimators of the change-points and the K-means clustering is used to group the drivers by their change-point estimates.

4.3 Simulation Study

A simulation study is conducted to check how the model performs.

One data set having two clusters from a NHPP with a piecewise constant intensity function is generated based on Klein and Roberts (1984). The censoring times of 40 drivers are sampled between 450 and 500 hours. The prior of the concentration parameter α_0 is $Gamma(0.01, 1)$. I run 30,000 iterations for the MCMC chain, discard the first 10,000 iterations, and use iterations obtained from every third observation after burn-in to do the inference. The |bias (%)| of a parameter τ is $|\widehat{\tau} - \tau|/\tau \times 100\%$.

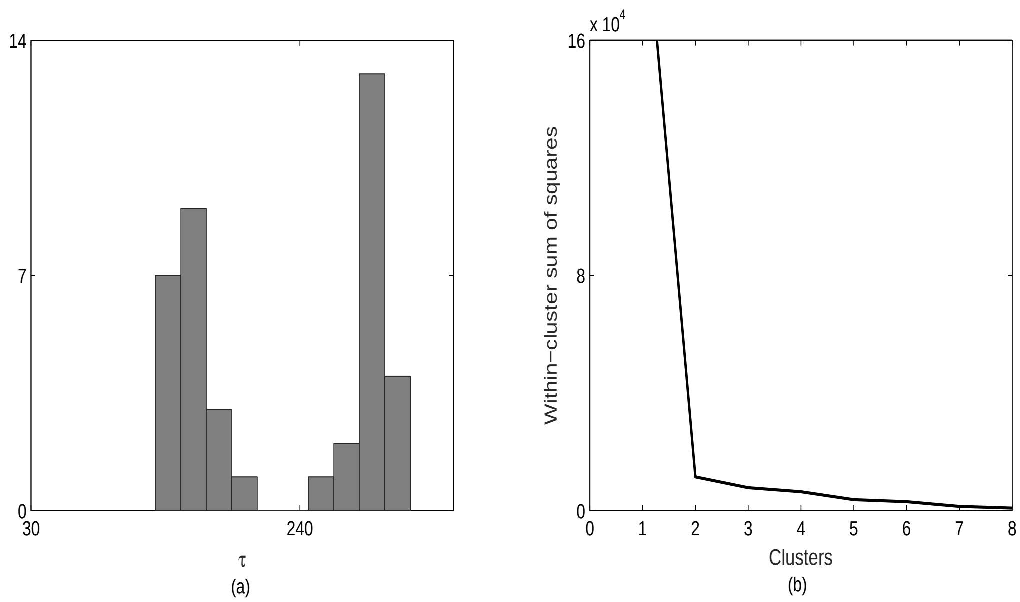

Figure 4.1: Simulated data: (a) histogram; (b) elbow plot.

Fig. 4.1 shows: (a) the histogram of the change-point estimates for the 40 drivers; (b) the elbow plot of the within-cluster sum of squares. From the two plots, it is clear that there are two clusters.

Table 4.1: The change-points of the simulated data

Parameter	True value	Average	SD	|Bias (%)|
μ_1	150	154.03	16.73	2.69
μ_2	300	296.95	17.97	1.02

All the drivers are assigned the correct membership. Table 4.1 shows the average and SD of the change-points for all drivers. Table 4.2 show the posterior means, standard deviations (SDs), absolute percentage bias and 95 % Highest Posterior Density (HPD) intervals of the intensity rates and concentration parameters.

Table 4.2: The intensity rates and the concentration parameter for the simulated data

Parameter	True value	Mean	SD	|Bias (%)|	95% HPD interval
λ_b	250	251.19	5.6	0.48	(240.18, 261.97)
λ_a	100	98.71	3.53	1.29	(91.80, 105.47)
α_0	-	18.44	3.6	-	(12.08, 25.44)

Fig. 4.2 shows the posteriors of the intensity rates and four drivers' change-points. Drivers 1 and 27 are in the first cluster and other two drivers are in the second cluster.

In summary, the DPMM detects the number of clusters correctly for the simulated data. The parameter estimations are satisfactory.

4.4 Application to the NTDS

I apply the model to the NTDS data and only consider the 38 teenagers with CNC events, because there is no information to estimate the change-points of the individuals without events.

The prior of the concentration parameter α_0 is $Gamma(0.01, 1)$. I run 80,000 iterations for the MCMC chain, discard the first 10,000 iterations, and use iterations obtained from every seventh observation after burn-in to do the inference.

Fig. 4.3 shows: (a) the histogram of the change-point estimates for the 38 drivers; (b)

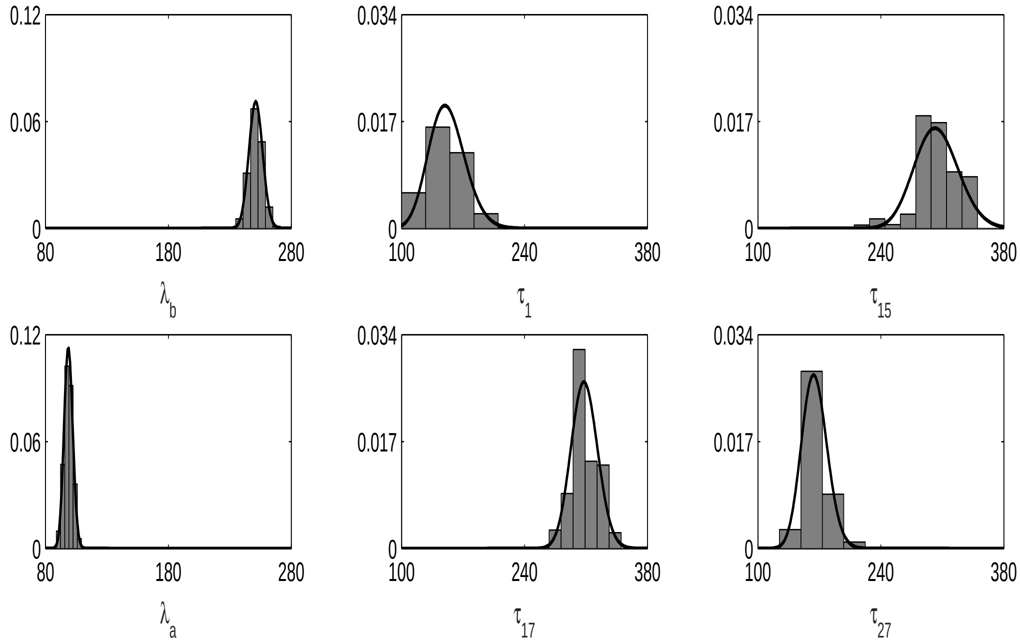

Figure 4.2: The posteriors of parameters for the simulated data

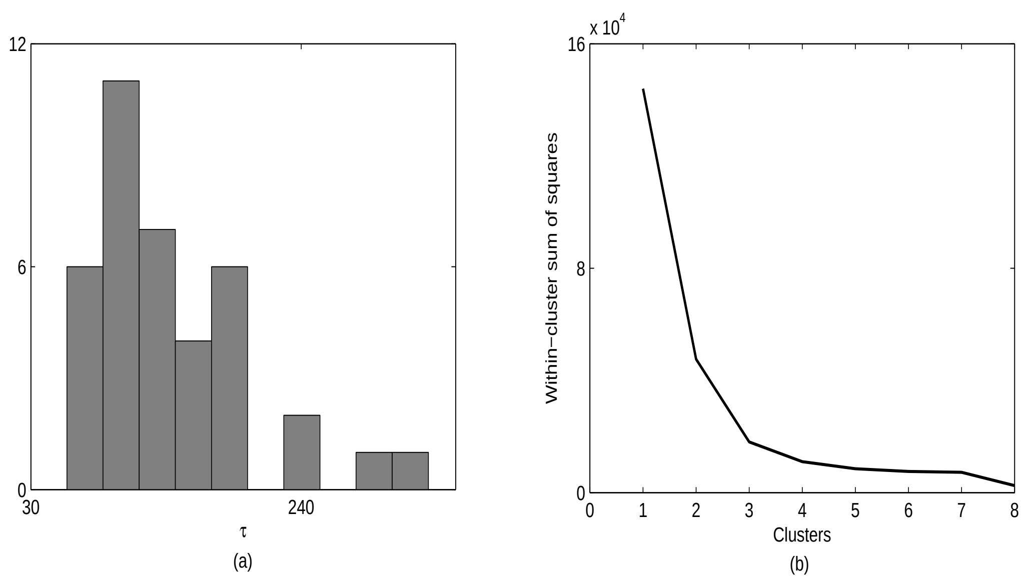

Figure 4.3: NTDS: (a) histogram of the change-points; (b) elbow plot.

the elbow plot of the within-cluster sum of squares. From these two plots, three clusters are chosen.

Table 4.3 and 4.4 show the posterior means, SDs, absolute percentage bias and 95 % HPD intervals of the parameters.

Table 4.3: The change-points of NTDS

Cluster	Parameter	Estimate	SD	Size
1	μ_1	96.31	16.38	22
2	μ_2	163.83	20.86	12
3	μ_3	279.19	45.23	4

Table 4.4: The intensity rates and concentration parameter of NTDS

Parameter	Mean	SD	95% HPD interval
λ_b	40.09	3.27	(33.80, 46.38)
λ_a	13.35	2.19	(9.21, 17.54)
α_0	18.69	3.62	(11.90, 25.72)

Based on the DPMM, three clusters exist among the teenage drivers. The change-points of the three clusters are 96.31, 163.83, and 279.19 hours. The intensity rate decreases from 40.09 CNC events per teenager per 1,000 hours to 13,35 after the change-point. 47.4% of the drivers fall in the first cluster, 39.5% in the second cluster and 13.1% in the third cluster.

Fig. 4.4 shows the posteriors of the intensity rates and four drivers' change-points. Drivers 15 is in the first cluster, and driver 1 is in the third cluster. Other two drivers are in the second cluster.

4.5 Conclusion and Discussion

Clusters exist among the teenagers with different patterns of risk change. It would be helpful to identify the underlying unknown structure.

In Chapter 3, I developed the Bayesian Hierarchical Finite Mixture Model to classify the teenagers and conducted model selection to choose the best number of clusters. Teh (2010)

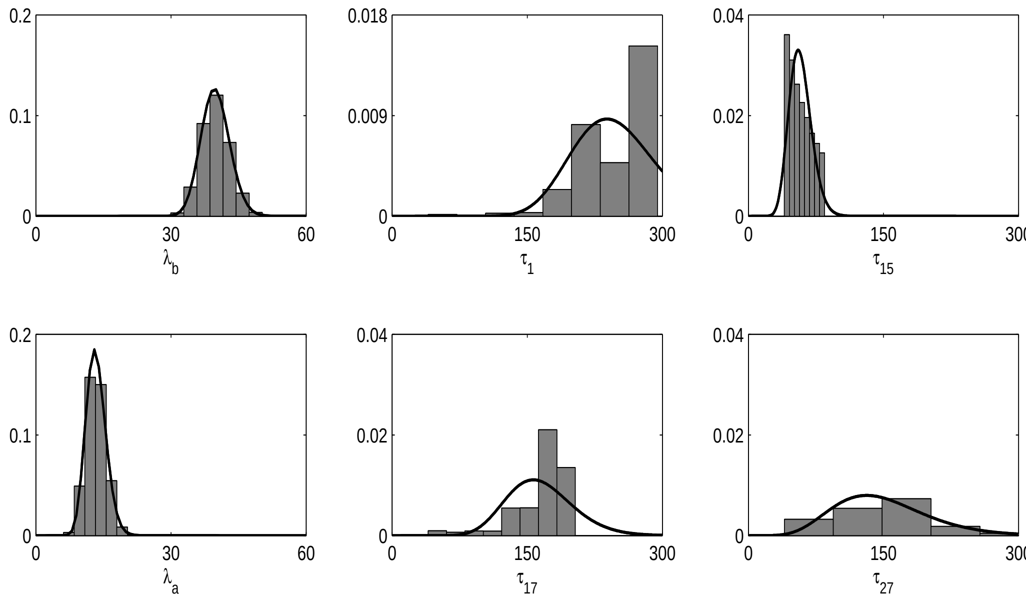

Figure 4.4: The posteriors of the parameters in NTDS

pointed out that an alternative was a Bayesian non-parametric approach. A non-parametric approach with a Dirichlet process prior is proposed to automatically detect the number of clusters in this chapter.

Based on the DPMM, three clusters exist among the teenage drivers. The change-points of the three clusters are 96.31, 163.83, and 279.19 hours. The intensity rate decreases from 40.09 CNC events per teenager per 1,000 hours to 13.35 after the change-point.

K-means clustering method is used after the MCMC in this chapter, which is straightforward to implement, but doesn't show the advantage of the automatic clustering property of DPMM. In the future, I would like to sample from the posterior of the number of clusters and use the posterior mode to determine the number of clusters. The posteriors of the allocation probabilities to different clusters of each subject will also be sampled. Then the subject is assigned to the cluster with the maximum membership probability. The posteriors of the parameters for each cluster will be sampled as well.

Chapter 5 CONCLUDING REMARKS

Crash rates among novice teenage drivers are high initially and decrease later. It is critical to identify the period during which novices are at greater risk so that parents and policy makers can provide more guidance to protect them and other drivers from their mistakes.

Previous research suggested that the change-point occurred at about six months of driving. However, the actual amount of driving time varies considerably during the first six months after licensure. Therefore, identifying the change-point of risk in driving hours provides critical information to improve young drivers' education, safety counter measures, and Graduated Driver Licensing regulations.

This dissertation develops three recurrent-event methods to detect the change-point of driving risk for novice teenage drivers. These three methods are under the assumption that the CNC event counts follow non-homogeneous Poisson process (NHPP) with constant piecewise intensity functions.

Chapter 2 proposes two recurrent-event change-point models to detect the time of change in driving risks, assuming that all the drivings share the identical change-points. It is shown that the change-points can only occur at the event times and the maximum likelihood estimators (MLE) are consistent. The MLE of the intensity rates are obtained first and the set of event times that maximize the profile likelihood are the MLE of the change-points. The NTDS application results indicate that crash and near-crash rate decreases significantly after 73 hours of independent driving after licensure. On the average, the intensity rate decreases from 32.31 CNC events per teenager per 10,000 hours to 20.84.

Chapter 3 proposes models to allow change-points to vary among drivers by a hierarchical Bayesian finite mixture model, considering that clusters exist among the teenagers. The drivers are assumed to fall in several clusters, and the subjects in the same cluster share the identical intensity rates and change-point. DIC is used to determine the number of clusters. The NTDS application shows three clusters exist among the teenagers: the change-points are 52.30, 108.99 and 150.20 hours of driving after first licensure correspondingly for the three clusters; the intensity rates increase for the first cluster while decrease for other two clusters; the change-point of the first cluster is the earliest and the average intensity rate is the highest.

Chapter 4 proposes a Dirichlet process mixture model allowing change-points to vary among drivers, and the change-points are assigned a Dirichlet process prior. A MCMC algorithm is developed to sample from the posterior distributions. Then the K-means clustering is used to group the subjects. Each subject has one unknown change-point ,and all the subjects share the identical intensity rates. Based on the DPMM, three clusters exist among the teenage drivers. The change-points of the three clusters are 96.31, 163.83, and 279.19 hours. The intensity rate decreases from 40.09 CNC events per teenager per 1,000 hours to 13.35 after the change-point.

The above three methods have similarities and close connections. The method in Chapter 2 by maximizing the likelihood is straightforward to implement and the NTDS application result is consistent with previous research. It also verifies the validity of the NHPP with constant piecewise intensity functions and provides reasonable estimates for the change-points and intensity rates, which can be the initial values and hyper parameters for other two methods. Other two methods extend the first one to accommodate the heterogeneity among teenagers. The Bayesian finite mixture model in Chapter 3 classifies the subjects in different groups, the individuals in the same cluster sharing the identical change-point and intensity rates. The NTDS application result is also consistent with previous research. The Dirichlet process mixture model in Chapter 4 clusters the teenagers automatically. This method is computationally expensive and the sampling is more complex.

The dissertation provides insight in teenagers' driving behaviour and will be critical to improve young drivers' education and the Graduated Driver Licensing regulations. The methodology has many applications across disciplines including the study of medical treatment effects, the hazard pattern of certain diseases, stock market analysis and quality control.

The future focus will be on incorporating covariates, predicting the change-points in driving risk of teenage drivers, and explore machine learning techniques.

Bibliography

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19:716–723.

Aldous, D. J. (1985). *Exchangeability and Related Topics*. Springer.

Andersen, P., Borgan, O., Gill, R., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer-Verlag, New York.

Antoniak, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian non-parametric problems. *The Annals of Statistics*, 2:1152–1174.

Aschar, J., Loibel, S., and Andrade, M. (2007). Interfailure data with constant hazard function in the presence of change-points. *REVSTAT*, 5:209–226.

Blackwell, D. and MacQueen, J. (1973). Ferguson distributions via pólya urn schemes. *The Annals of Statistics*, 1:353–355.

Blei, D. M. and Jordan, M. I. (2006). Variational inference for Dirichlet process mixtures. *Bayesian Analysis*, 1(1):121–143 (electronic).

Burnham, K. P. and Anderson, D. (2002). *Model Selection and Multi-Model Inference: a Practical Information-Theoretic Approach*. Springer-Verlag, New York.

Bush, C. A. and MacEachern, S. N. (1996). A semiparametric Bayesian model for randomised block designs. *Biometrika*, 83(2):275–285.

Celeux, G., Hurn, M., and Robert, C. P. (2000). Computational and inferential difficulties with mixture posterior distributions. *Journal of the American Statistical Association*, 95:957–970.

Centers for Disease Control and Prevention (2008). Web-based injury statistics query and reporting system (wisqars): leading causes of death peports 19992005. `http://webappa.cdc.gov/sasweb/ncipc/leadcaus10.html`. [Online].

Cook, R. J. and Lawless, J. F. (2007). *The Statistical Analysis of Recurrent Events*. Statistics for biology and health. Springer.

Deery, H. and Fildes, B. (1999). Young novice driver subtypes: relationship to high-risk behavior, traffic accident record, and simulator driving performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 41:628–43.

Diebolt, J. and Robert, C. P. (1994). Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society B*, 56:363–375.

Dingus, T., Klauer, S., Neale, V., Petersen, A., Lee, S., Sudweeks, J., and et al. (2006). The 100-car naturalistic driving study: phase ii results of the 100-car field experiment. Technical Report DOT-HS-810-593.

Dupuy, J.-F. (2006). Estimation in a change-point hazard regression model. *Statistics and Probability Letters*, 76:182–190.

Escobar, M. (1994). Estimating normal means with a Dirichlet Process prior. *Journal of the American Statistical Association*, 89:268–277.

Escobar, M. and West, M. (1995). Bayesian density estimation and inference using mixtures. *The Annals of Statistics*, 90:577–588.

Ferguson, T. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1:209–230.

Ferguson, T. (1974). Prior distributions on spaces of probability measures. *The Annals of Statistics*, 1:615–629.

Ferguson, T. S. (1983). Bayesian density estimation by mixtures of normal distributions. *Recent Advances in Statistics*, 24:287–302.

Field, C. A. and Welsh, A. H. (2007). Bootstrapping clustered data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69:369–390.

Fitch, G. M., Soccolich, S. A., Guo, F., McClafferty, J., Fang, Y., Olson, R. L., Perez, M. A., Hanowski, R. J., Hankey, J. M., and Dingus, T. A. (2013). The impact of hand-held and hands-free cell phone use on driving performance and safety-critical event risk. Technical Report DOT-HS-811-757.

Frobish, D. and Ebrahimi, N. (2009). Parametric estimation of change-points for actual event data in recurrent events models. *Computational Statistics and Data Analysis*, 53:671–682.

Gelman, A. and Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7:457–511.

Ghosh, J., Joshi, S., and C., M. (1993). A Bayesian approach to the estimation of the change-point in a hazard rate. *Advances in Reliability*, pages 141–170.

Guo, F. and Fang, Y. (2013). Individual driver risk assessment using naturalistic driving data. *Accident Analysis & Prevention*, 61:3–9.

Guo, F., Fang, Y., and Antin, J. F. (2015a). Evaluation of older driver fitness-to-drive metrics and driving risk using naturalistic driving study data. Technical Report 15-UM-036.

Guo, F., Fang, Y., and Antin, J. F. (2015b). Older driver fitness-to-drive evaluation using naturalistic driving data. *Journal of Safety Research*.

Guo, F., Klauer, S., Hankey, J., and Dingus, T. (2010). Using near-crashes as a crash surrogate for naturalistic driving studies. *Journal of the Transportation Research Board*, 2147:66–74.

Guo, F., Simons-Morton, B. G., Klauer, S. E., Ouimet, M. C., Dingus, T. A., and Lee, S. E. (2013). Variability in crash and near-crash risk among novice teenage drivers: a naturalistic study. *The Journal of Pediatrics*, 163:1670–1676.

Heller, K. A. and Ghahramani, Z. (2005). Bayesian hierarchical clustering. In *Proceedings of the 22Nd International Conference on Machine Learning*, ICML '05.

Hjort, N. L., Homes, C., Muller, P., and Walker, S. G. (2010). *Bayesian Nonparametrics*. Cambridge University Press.

Jackson, J. C., Albert, P. S., Zhang, Z., and Simons-Morton, B. (2013). Ordinal latent variable models and their application in the study of newly licensed teenage drivers. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 62:435–450.

Karasoy, D. and Kadilar, C. (2007). A new bayes estimate of change-point in the hazard function. *Computational Statistics and Data Analysis*, 51:2993–3001.

Keating, D. (1978). A search for social intelligence. *Journal of Educational Psychology*, 70:218–223.

Klauer, S., Dingus, T. A., Neale, V. L., Sudweeks, J., and Ramsey, D. J. (2009). Comparing real-world behaviors of drivers with high vs. low rates of crashes and near-crashes. Technical Report DOT-HS-811-091, National Highway Traffic Safety Administration, Washington, DC.

Klauer, S., Guo, F., Sudweeks, J., and Dingus, T. (2010). An analysis of driver inattention using a case-crossover approach on 100-car data. Technical Report DOT-HS-811-334, National Highway Traffic Safety Administration, Washington, DC.

Klauer, S. G., Guo, F., Simons-Morton, B. G., Ouimet, M. C., Lee, S. E., and Dingus, T. A. (2014). Distracted driving and risk of road crashes among novice and experienced drivers. *The New England Journal of Medicine*, 370:54–59.

Klein, R. W. and Roberts, S. D. (1984). A time-varying poisson arrival process generator. *Simulation*, 43:193–5.

Kotz, S., Balakrishnan, N., and Johnson, N. L. (2000). *Continuous Multivariate Distributions*, volume 1. Wiley.

Lawless, J. and Zhan, M. (1998). Analysis of interval-grouped recurrent-event data using piecewise constant rate functions. *Canadian Journal of Statistics*, 26:549–565.

Lazarsfeld, P. and Henry, N. (1968). *Latent Structure Analysis*. Wiley.

Lee, S., Simons-Morton, B., Klauer, S., Ouimet, M., and Dingus, T. (2011). Naturalistic assessment of novice teenage crash experience. *Accident Analysis & Prevention*, 43:1472–9.

Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation (2nd ed.)*. Springer, New York.

Li, Q., Guo, F., Klauer, S., and Simons-Monton, B. (2015). Recurrent-event change-point models for novice teenage driving risk evaluation. submitted.

Liu, J. S. (2003). The collapsed gibbs sampler in bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association*, 89:958–966.

Loader, C. R. (1991). Inference for a hazard rate change-point. *Biometrika*, 78:749–757.

MacEachern, S. N. and Müller, P. (1998). Estimating mixture of Dirichlet Process models. *Journal of Computational and Graphical Statistics*, 7:223–238.

Matthews, D. and Farewell, V. (1982). On testing for a constant hazard against a change-point alternative. *Biometrics*, 38:463–468.

Mayhew, D., Simpson, H., and Pak, A. (2003). Changes in collision rates among novice drivers during the first months of driving. *Accident Analysis & Prevention*, 35:683–91.

McCutcheon, A. (1987). *Latent Class Analysis.* Beverly Hills: Sage Publications.

McLachlan, G. and Peel, D. (2000). *Finite Mixture Models.* Wiley Series in Probability and Statistics: Applied Probability and Statistics. Wiley-Interscience, New York.

Minka, T. and Ghahramani, Z. (2003). Expectation and propagation for infinite mixtures. In *Workshop on Nonparametric Bayesian Methods and Infinite Models*, NIPS 03.

Müller, H. G. and Wang, J.-L. (1994). Change-point models for hazard functions. *IMS Lecture Notes - Monograph Series*, 23.

Murphy, K. P. (2012). *Machine Learning: a Probabilistic Perspective.* MIT press.

Musselwhite, C. (2006). Attitudes towards vehicle driving behavior: categorizing and contextualizing risk. *Accident Analysis & Prevention*, 38:324–34.

National Highway Traffic Safety Administration (2000). Traffic safety facts 2000: young drivers. Technical Report DOT-HS-809-336, National Center for Statistics & Analysis, Washington DC.

Neal, R. M. (1992). Bayesian mixture modeling. *In Proceedings of the Workshop on Maximum Entropy and Bayesian Methods of Statistical Analysis*, 11:197–211.

Neal, R. M. (1998). Morkov chain sampling methods for Dirichlet process mixture models. Technical Report 9815, Department of Statistics, University of Toronto.

Neal, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 9:249–265.

Neal, R. M. (2003). Slice sampling. *The Annals of Statistics*, 31:705–767.

Nguyen, H., Rogers, G., and Walker, E. (1984). Estimation in change-point hazard rate models. *Biometrika*, 71:299–304.

Nobile, A. (2007). Bayesian finite mixtures: a note on prior specification and posterior computation. Technical Report 05-3, Department of Statistics, University of Glasgow.

Orbanz, P. (2010). Construction of nonparametric Bayesian models from parametric Bayes equations. *In Advances in Neural Information Processing Systems.*

Ouimet, M. C., Brown, T. G., Guo, F., Klauer, S. G., Simons-Morton, B. G., Fang, Y., Lee, S. E., Gianoulakis, C., and Dingus, T. A. (2014). Higher crash and near-crash rates in teenaged drivers with lower cortisol response: an 18-month longitudinal, naturalistic study. *JAMA Pediatrics*, (6):517–522.

Pons, O. (2002). Estimation in a cox regression model with a change-point at an unknown time. *Statistics*, 36:101–124.

Richardson, S. and Green, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society B*, 59:731–792.

Rolls, G., Hall, R., Ingham, R., and McDonald, M. (1991). *Accident Risk and Behavioral Patterns of Younger Drivers.* AA Foundation for Road Safety Research, Fanum House, Basingstoke, Hampshire, UK.

Ross, S. M. (2006). *Simulation (4th ed.).* Academic Press.

Sethuraman, J. (1994). A constructive definition of the Dirichlet process prior. *Statistica Sinica*, 2:639–650.

Simons-Morton, B. G., Guo, F., Klauer, S. G., Ehsani, J. P., and Pradhan, A. K. (2014). Keep your eyes on the road: young driver crash risk increases according to duration of distraction. *Journal of Adolescent Health*, 54:S61–S67.

Simons-Morton, B. G., Ouimet, M. C., Zhang, Z., Lee, S. E., Klauer, S. E., Wang, J., Albert, P. S., and Dingus, T. A. (2011a). Crash and risky driving involvement among novice adolescent drivers and their parents. *American Journal of Public Health*, 101:2362–2367.

Simons-Morton, B. G., Ouimet, M. C., Zhang, Z., Lee, S. E., Klauer, S. E., Wang, J., Chen, R., Albert, P. S., and Dingus, T. A. (2011b). The effect of passengers and risk-taking friends on risky driving and crashes/near crashes among novice teenagers. *Journal of Adolescent Health*, 49:587–593.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B*, 64:583–639.

Stephens, M. (2000). Bayesian analysis of mixture models with an unknown number of componentsan alternative to reversible jump methods. *The Annals of Statististics*, 28:40–74.

Teh, Y. W. (2010). *Encyclopedia of Machine Learning*, chapter Dirichlet Process, pages 280–287. Springer US.

Teh, Y. W., Jordan, M. I., Beal, M. J., and Blei, D. M. (2006). Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 101:1566–1581.

Thompson, W. (1988). *Point Process Models with Applications to Safety and Reliability*. Chapman and Hall, UK.

Ulleberg, P. (2001). Personality subtypes of young drivers. relationship to risk-taking preferences, accident involvement, and response to a traffic safety campaign. *Transportation Research Part F: Traffic Psychology and Behaviour*, 4:279–97.

Van Der Vaart, A. (1998). *Asymptotic Statistics*. Cambridge University Press, UK.

Walker, S. and Damien, P. (1998). Sampling methods for Bayesian nonparametric inference involving stochastic processes. In *Practical Nonparametric and Semiparametric Bayesian Statistics*, pages 243–254. Springer.

Walker, S. G., Damien, P., Laud, P. W., and Smith, A. F. M. (1999). Bayesian nonparametric inference for random distributions and related functions. *Journal of the Royal Statistical Society B*, 61:485–527.

Wang, L. and Dunson, D. B. (2011). Fast Bayesian inference in Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 20:196–216.

West, M., Muller, P., and Escobar, M. (1994). Hierarchical priors and mixture models with applications in regression and density estimation. *Aspects of Uncertainty*, pages 363–386.

West, R. and Odgen, R. (1997). Continuous-time estimation of a change-point in a poisson process. *Journal of Statistical Computational Simulation*, 56:293–302.

Williams, A. (2003). Teenage drivers: patterns of risk. *Journal of Safety Research*, 34:5–15.

Wu, C., Zhao, L., and Wu, Y. (2003). Estimation in change-point hazard function models. *Statistics and Probability Letters*, 63:41–48.

Yao, Y. C. (1986). Maximum likelihood estimation in hazard rate models with a change-point. *Communications in Statistics - Theory and Methods*, 15:2455–2466.